The Neural Bases of Individual Differences in Complex Cognition and Behavior

By

Adam Skyler Eichenbaum

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Neuroscience

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Mark D'Esposito, Chair
Professor Rich Ivry
Professor Anne Collins
Professor Dennis Levy
Professor C. Shawn Green

Fall 2020

Abstract

The Neural Bases of Individual Differences in Complex Cognition and Behavior

by

Adam Skyler Eichenbaum

Doctor of Philosophy in Neuroscience

University of California, Berkeley

Professor Mark D'Esposito, Chair

Everyday human cognition and behavior is accomplished via the coordinated efforts of numerous complex processes. From retaining information in memory to maintaining long-term personal goals, human behavior is multifaceted. In light of this complex nature, there exists massive variability from human to human in their ability to perform and implement these behaviors. In this dissertation, I present three experiments that elucidate the neural processes underlying this variability. Each experiment involved the collection and analysis of functional magnetic resonance imaging (fMRI) data. In one of these experiments, subjects performed a cognitive task during fMRI, while the other two assessed the relationship between baseline neural activity and complex cognition.

The first chapter of this dissertation uses behavioral modeling and fMRI to assess individual differences in the ability to learn and subsequently implement a hierarchically-structured rule set. I show that humans are capable of learning and implementing the complex hierarchical rule and that neural activity across multiple networks supports this ability. First, I find that brain regions across frontal and parietal cortices support the initial discovery of the hierarchical rule. Next, activity across a cingulo-opercular network of brain regions supports the generalization of this knowledge to novel settings.

I present evidence in the second chapter that individual differences in cognition and behavior are not only predicted by the patterns of coordinated neural activity across the entire brain, but that novel temporally-varying analysis approaches provide additional predictive power unobtainable with previous approaches. Here, I analyzed fMRI data collected while subjects performed no explicit task, a procedure referred to as "resting-state" fMRI. Subjects also completed a set of cognitive computer tasks that measured complex cognitive abilities such as working memory and cognitive control. By applying both traditional time-invariant and novel time-varying graph theoretical analyses to the resting-state fMRI data, I was able to predict individual differences in the cognitive abilities measured outside the MRI scanner. Moreover, the novel time-varying analysis revealed relationships to behavior that better captured task-specific behavioral variability.

The final chapter examines the ability of resting-state graph theoretical approaches to predict cognitive abilities related to attention and inhibitory control. Moreover, neural measures were interrogated alongside measures of human physiological functioning. Here, I find that all attentional and inhibitory abilities are accurately predicted across all neural and physiological measures. Specifically, neural measures indicative of brain-wide activity patterns predicted attentional accuracy and global inhibition ability, while neural measures reflecting activity patterns of particular neural networks additionally predicted attentional reactivity. Further, measures of physiological functioning were able to predict individual aspects of inhibitory control.

Together, the three experiments presented here contribute to our knowledge of how neural activity patterns ultimately beget complex cognition and behavior. Using multiple complimentary analytical approaches, I find evidence for the role of multiple neural networks in explaining the differences in cognitive abilities across the healthy human population.

Table of Contents

Acknowledgements

First, I need to bring attention to the numerous people in lab who are responsible, in no small part, for helping me get to where I am today. Let us begin with the D'Esposito lab. Without the smallest of doubts in my mind, I want to, and need to, thank Dr. Elizabeth Lorenc for her (truly) unending support and guidance over the years. No matter the question, no matter the time, Elizabeth stopped everything she was doing to help me. Every. Single. Time. She will deny this. Don't believe her. I would like to additionally thank my office mates across the years (Maxwell Bertolero, Jacob Miller, Daniel Lurie, Daniel Toker). Special recognition needs to be made for Jason Scimeca and Ioannis Pappas. Both individuals took me under their wing throughout my 5 years at Berkeley and served as extraordinary mentors. Although meetings were planned with a conservative frequency of once a week, both quickly learned that face-to-face sessions would greatly exceed that unreasonably low initial rate. I look back fondly on every meeting in the game room, solarium, and in-office rapid drive-bys (when they foolishly revealed themselves to be physically in lab). I would like to formally acknowledge and thank my co-authors on Chapter 2 (Jason Scimeca) and Chapter 3 (Ioannis Papper, Daniel Lurie, Jessica Cohen).

Beyond the expansive boundaries of the D'Esposito lab, atop the Helen Wills hierarchy, is Candace Groskreutz. She is so much more than just the Graduate Program Manager and deserves pages upon pages of personal and professional recognition. Yes, she is great at her job, but that barely begins to illustrate her value. Candace made it feel like you were the only graduate student in the program. Her personal, and unreservedly sincere, compassion for and consideration of even the smallest of matters speaks volumes to her character. Lacking even the slightest embellishment, and with earnestness that is seldom expressed, I want to say thank you for everything that you have done for me.

Leaving the University of California behind, I want to bring attention to C. Shawn Green. I owe so many thanks to you for your focused and personalized mentorship during my time at the University of Wisconsin (Go Badgers!), which even continued after I moved 1,700 miles away. Your friendship and care are a constant source of inspiration and personal pride.

Without a doubt though, I want to thank my family, whose selfless support proved invaluable. First, I want to thank my brother Alex, whose regular visits and eventual laying down of roots in the Bay Area brought me endless joy. From facetime-ing "Playclock at 5… the pass is INTERCEPTED" while unofficially interviewing for grad school at my advisor's house, to watching a 440-minute 18-inning World Series Game 3, to never, ever boiling your ribs, it was great to know we were in the good old days while they were actually happening.

Last, I want to say how much I love my parents, Karen and Howard. My father would be beaming to see me graduate; I know just how lucky I am to have to have had him in my life for as long as I did. My parents' impassioned interest and genuine belief in my ability to achieve anything I wanted continues to encourage me. I cannot thank them enough for everything they have done for me.

CHAPTER 1

1.1 Introduction

Once thought to be governed by humble stimulus-response conditioning, human behavior is now known to be the result of a multitude of complex processes that dynamically interact to produce everyday human behaviors. From learning how to adapt to a new social setting, to maintaining relevant information in memory, and attending to various streams of information in order to achieve short- and long-term goals, human behavior is complex and highly variegated. Alongside this complexity comes variability across humans in their ability to perform these behavioral functions. The goal of this dissertation is to elucidate the neural mechanisms responsible for this observed variability in complex human behavior.

In Chapter 2 I focus on the neural mechanisms associated with the learning and generalization of complex rules. Although seemingly mundane, the simple act of visiting a new coffee shop requires generalizing knowledge from previous experiences. You must determine where the line forms, what the menu consists of, and how to place your order and retrieve your drink. Humans are uniquely capable of adapting to novel scenarios such as these, evidenced by the fact that such an act itself seems unremarkable. However, the processes involved in learning new information and generalizing it to novel scenarios are highly complex, and the neural mechanisms underlying this ability are only beginning to be understood. Regions along the right lateral frontal surface of the brain have been associated with the learning and implementation of complex, hierarchically structured rules such as these (Badre & D'Esposito, 2007; Badre & Nee, 2018). In addition, regions comprising a "cingulo-opercular" network that spans frontal cortex, insular cortex, and the subcortex have been associated with the maintenance of structured task sets that coordinate a series of goal-relevant behaviors (Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008; Dosenbach et al., 2006). It remains uncertain, though, how the brain supports the generalization of this knowledge to novel settings. Using fMRI and behavioral modeling, I discovered that a network of regions spanning both frontal and parietal cortices work together to support the initial discovery of a hidden hierarchical rule that governs how subjects should respond to all incoming stimuli. In addition, I found that individual differences in the ability to generalize this rule to novel settings was related to the variability in the activity throughout a cingulo-opercular network of brain regions. These results demonstrate that the uniquely human ability of learning and adaptation relies on a dual-network architecture of brain areas spanning multiple cortical regions.

In Chapter 3 I investigate how distributed patterns of neural activity measured during a task-free baseline resting state predict individual differences in working memory and cognitive control. Working memory is a core cognitive function and refers to the active short-term retention of information that is no longer present in the environment (D'Esposito & Postle, 2015; Fuster, 1973). In addition, cognitive control is a foundational human ability and broadly refers to the collective ability of maintaining and processing selective information from multiple streams in order to perform specific behaviors (M. M. Botvinick, Braver, Barch, Carter, & Cohen, 2001). Cognitive abilities that tap these crucial functions have been interrogated using resting-state fMRI analyses in the past with great success (Biswal, Yetkin, Haughton, & Hyde, 1995; Cohen & D'Esposito, 2016; Stevens, Tappon, Garg, & Fair, 2012). These resting-state fMRI analyses

have typically analyzed neural activity as a static, time-invariant property, however recent findings have shown that considering the data from a time-varying perspective may allow for additional insight (Cohen, 2018; Sadaghiani, Poline, Kleinschmidt, & D'Esposito, 2015; Thompson, 2018; Vidaurre, Smith, & Woolrich, 2017). Specifically, the time-varying nature of these analyzes may be better situated to capture variability in these core cognitive abilities. By contrasting time-invariant against time-varying resting-state fMRI analysis approaches, I discovered that information carried in time-varying resting-state neural activity patterns better captured behavioral differences compared to time-invariant methods. Specifically, time-invariant analyses tracked trait-level attributes such as general multi-task cognitive abilities and personality traits, while time-varying analyses additionally tracked individual differences in the variability of working memory reaction times. These results provide evidence that time-varying fluctuations in rest-state neural activity patterns carry meaningful information related to the variability of complex human behaviors such as working memory.

In Chapter 4 I investigate individual differences in core attentional and inhibitory cognitive abilities using resting-state fMRI alongside the analysis of measures of physiological functioning. Building off the previous analysis of resting-state fMRI data, I incorporate the collection of measures of cardiac and respiratory functioning in order to predict individual differences in cognitive abilities. I found that although neural measures derived from resting-state fMRI and physiological measures tracked performance variability across multiple attention-based cognitive tasks, unique prediction differences existed within and between the two measurement modalities. First, neural measures indexing brain-wide activity patterns tracked measures of general accuracy and global inhibition ability, while activity patterns within specific neural networks additionally tracked measures of general reaction time. Measures of physiological functioning, in addition to tracking measures of general reaction time and global inhibition ability, were uniquely capable of teasing apart differences related to inhibition reaction time and inhibition accuracy. These results suggest that predictive models of complex human behavior can benefit by the inclusion of extra-neuronal biologically-relevant features.

CHAPTER 2

Dissociable neural systems support the learning and transfer of hierarchical control structure

2.1 Abstract

Humans can draw insight from previous experiences in order to quickly adapt to novel environments that share a common underlying structure. Here we combine functional imaging and computational modeling to identify the neural systems that support the discovery and transfer of hierarchical task structure. Human subjects (male and female) completed multiple blocks of a reinforcement learning task that contained a global hierarchical structure governing stimulus-response action mapping. First, behavioral and computational evidence showed that humans successfully discover and transfer the hierarchical rule structure embedded within the task. Next, analysis of fMRI BOLD data revealed activity across a frontal-parietal network that was specifically associated with the discovery of this embedded structure. Finally, activity throughout a cingulo-opercular network supported the transfer and implementation of this discovered structure. Together, these results reveal a division of labor in which dissociable neural systems support the learning and transfer of abstract control structures.

2.2 Introduction

Whether it is learning how to drive a new car or interacting with an unfamiliar social group, humans show remarkable adaptability inferring the correct action given minimal information. Such learning usually occurs via trial and error where feedback works to guide future behavior. These problem-solving approaches are routinely accelerated by generalizing previous knowledge (Woodworth & Thorndike, 1901). When simple stimulus-response mappings are learned in experimental settings, responses learned in one context can be directly transferred to a subsequent context, leading to an immediately observable benefit (Behrens et al., 2007; Collins et al., 2014; Collins & Frank, 2016). Although humans can encounter scenarios such as these (e.g., opening computer applications on a Windows vs. Apple operating system), humans also encounter settings where this approach leads to failure (e.g. starting computer programs on Windows/Apple vs. Linux). In these cases, it is advantageous to instead leverage prior knowledge to guide the learning of the correct behavior, a process known as "learning to learn" (Bavelier et al., 2012; Botvinick et al., 2019; Harlow, 1949; Kemp et al., 2010). While the behavioral and neurobiological underpinnings of more direct types of transfer have been relatively well-characterized (Collins et al., 2014; Collins & Frank, 2016), the neural systems and mechanisms underlying this more abstract form of transfer remain poorly understood.

Everyday experiences are often structured hierarchically where actions and experiences are influenced by superordinate contexts and rules. For example, when traveling away from home it is common to pack a bag with clothes and overnight necessities. However, the rule that

restricts packing small-volume liquids is only relevant in certain contexts: when traveling by airplane, not by car. By grouping these sets of behaviors and experiences hierarchically, one is able to easily generalize rules from one context to another, and even to contexts that have not yet been personally experienced. One way in which learned hierarchical structures may be generalized to novel contexts is the creation of task sets or task structures that span across related contexts regardless of low-level features (Collins & Frank, 2013).

Although the combination of task sets and hierarchical processing provides a natural candidate solution for how learned hierarchical structure is generalized, the neural basis of these cognitive processes has typically been studied in isolation. Growing neurobiological and computational evidence suggests that the frontal cortex facilitates hierarchically-structured behavior (Badre & D'Esposito, 2007; Badre et al., 2010; Collins & Frank, 2013; Frank & Badre, 2012; Koechlin, 2003; Nee & D'Esposito, 2016; Wang et al., 2018). Specifically, left lateral frontal cortex is organized along a rostrocaudal gradient wherein more rostral regions support the learning and execution of increasingly higher-order hierarchically-structured rules. It remains undetermined whether these regions, likely those more rostrally, additionally support the transfer of learned structure (Badre & Nee, 2018). In addition, processing of task sets has generally been related to activity in frontal cortex, as well as to a distributed "cingulo-opercular" network of regions (Dosenbach et al., 2008; Sakai, 2008). As generalization of hierarchical knowledge involves the integration of information across multiple sources, it is likely that a network of regions spanning beyond frontal cortex will be involved.

To investigate the discovery and transfer of abstract hierarchical structure, we designed a hierarchical reinforcement learning task that promotes the creation and transfer of a superordinate structure (Figure 2.1). Critically, although each block contained entirely new stimulus features, a global 2nd-order hierarchical rule remained. Therefore, successful performance of a previous block conveyed no immediate advantage on subsequent blocks. However, knowledge of the correct hierarchical structure instead facilitated a more rapid learning of the correct response mappings. We leveraged converging computational modeling approaches to confirm (1) when subjects first discovered the global hierarchical structure, and (2) that rapid learning occurred thereafter, indicating transfer of learned structure. Lastly, we used fMRI to investigate the left lateral frontal regions along the predefined rostrocaudal gradient, as well as broader neural systems, that support these two processes.

2.3 Materials and Methods

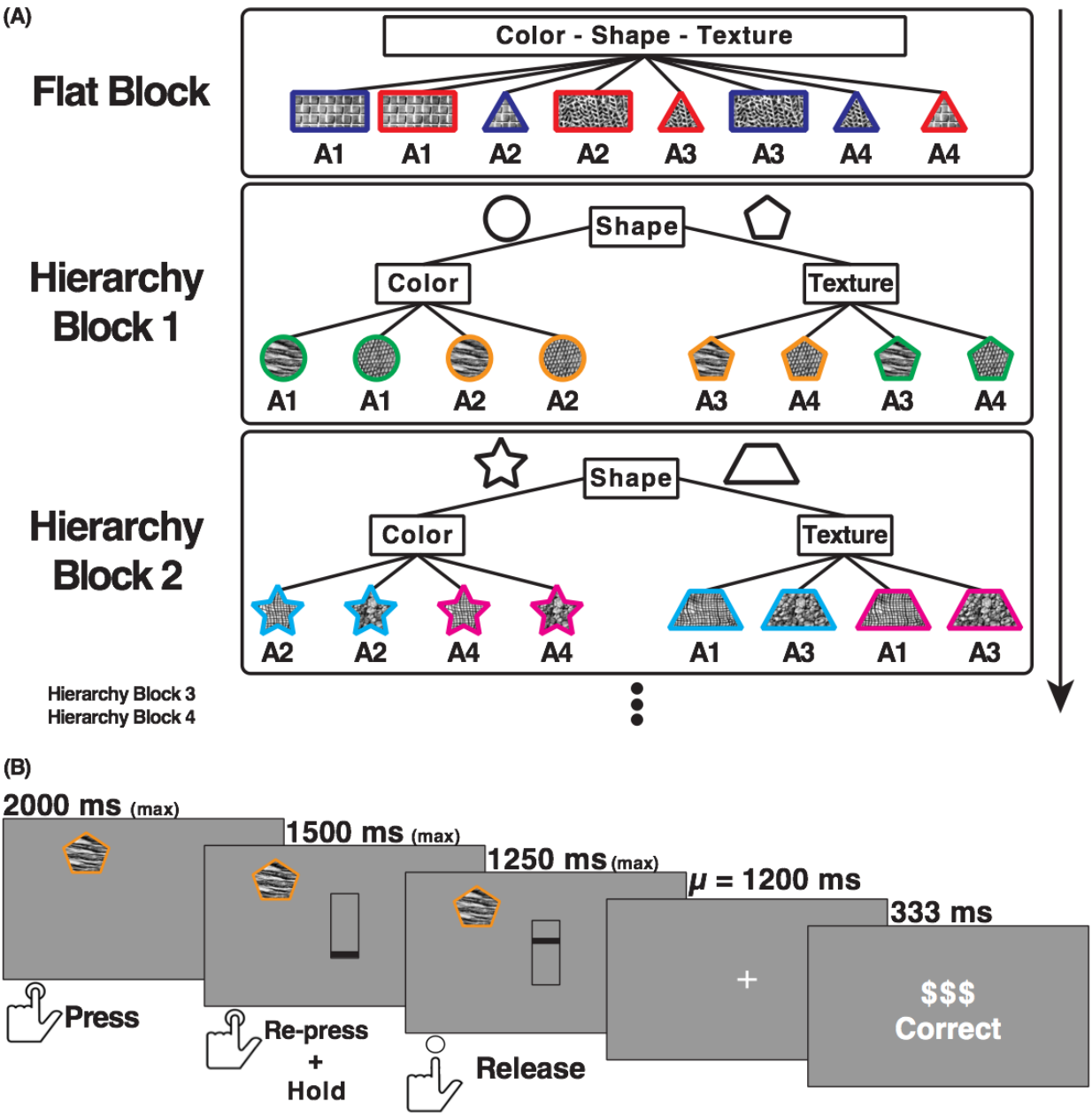


Figure 2.1 Schematic Depiction of Experimental Logic and Trial Sequence. (A) Schematic of task design showing example stimulus-to-action mappings. Subjects completed five blocks in total throughout the experiment. The stimuli in each block varied along three dimensions: shape, color, and texture. Each block contained two stimulus features for each dimension (e.g. two shapes) and the specific features changed for each block. The first block contained a flat policy structure such that the mapping between stimuli and actions (A1, A2, etc.) was randomly assigned. The remaining four blocks all shared the same global 2nd-order policy structure: the shape of the stimulus indicated whether first-order rules were determined by color or texture on the current trial. In the example shown for hierarchical block 1, a circular stimulus indicated that color determined the correct action (i.e., green pairs with A1, orange pairs with A2). Hierarchical blocks included an irrelevant fourth dimension (stimulus position on screen) that is not shown here. (B) Schematic of trial design. Trials began with stimulus presentation, after which subjects had up to 2s to respond by pressing one of four buttons mapped to their right index, middle, ring, and pinky fingers. Subjects then indicated their confidence in their answer by positioning a black bar along the screen in a one-shot manner. Subjects received auditory and visual feedback following a jittered ISI.

Human Subject Details

Thirty-two healthy right-handed subjects (range: 18 – 29 years; mean = 19.63; SD = 2.54; 20 females) with normal or corrected-to-normal vision participated in the study at the University of California – Berkeley. Target sample size was based on prior relevant literature (Badre et al., 2010; Collins & Frank, 2016; Nee & D'Esposito 2016). Eight subjects were excluded from all behavioral analyses (four subjects failed to complete the entire session, two subjects did not follow the instructions, and two subjects exhibited sub-threshold behavioral performance (no above-chance performance in any hierarchical block, i.e., state-space model outcomes of the distribution around the probability to produce a correct response always included the chance-level performance value)). Five additional subjects were excluded from all fMRI analyses (one subject due to above-threshold in-scanner motion (>2.5mm in X, Y, or Z across all blocks), one subject for atypical anatomical data and three subjects due to scanner image reconstruction failures).

All behavioral analyses presented here include data from the 24 subjects for whom we obtained a complete behavioral dataset (range: 18 – 24 years; mean = 19.25; SD = 1.75; 16 females). All fMRI analyses presented here include data from the 19 subjects for whom we obtained a complete behavioral and fMRI dataset (range: 18 – 24 years; mean = 19.26; SD = 1.88; 13 females). Behavioral analyses restricted to these 19 subjects show the same results as the 24-subject group. All research protocols were approved by the Committee for Protection of Human Subjects at the University of California, Berkeley. Informed and written consent was obtained from all subjects prior to participation.

Experimental Design and Statistical Analyses

Task Design

In the current experiment, we designed a reinforcement learning task (inspired by Badre et al., 2010) that required learning multiple distinct 2^{nd}-order hierarchical rules (hereafter referred to as 2^{nd}-order policy) that shared a global hierarchical policy structure (hierarchical blocks). Specifically, a 2^{nd}-order hierarchical policy determined that the shape of the stimulus cued 1^{st}-order rules defined by other stimulus dimensions (e.g. if the stimulus is a square, perform action 1 for red squares, and action 2 for blue squares, however if the stimulus is a circle, perform action 3 for striped circles, and action 4 for checkered circles, regardless of other stimulus features). Thus, subjects who learn the block-specific hierarchical policy in successive blocks can discover the existence of the global hierarchical structure. By transferring their knowledge of the global hierarchical structure to subsequent blocks, subjects can more rapidly learn the block-specific hierarchical policy.

Subjects completed one block containing a rule set in which there was no higher-order structure (flat block) and four hierarchical blocks while inside the scanner (Figure 2.1). Subjects viewed stimuli that varied along three or four dimensions: shape, color, black-and-white image pattern (referred to as "texture"), and stimulus position on screen (hierarchical blocks only) (Figure 2.1A). For each block, stimulus dimensions could vary between two features (e.g. Color: red/blue, Shape: square/circle, etc.), resulting in 8 unique stimuli in the flat block and 16 unique

stimuli in each hierarchical block. All blocks contained unique features, and thus subjects had to learn entirely new stimulus-response mappings for each block. We assigned stimulus features to blocks by random assignment.

Stimuli

Stimuli were generated using PsychoPy (Peirce, 2007, 2008). Colors included red, green, blue, yellow, magenta, cyan, white, maroon, black, and orange. Shapes included a circle, square, rectangle, triangle, pentagon, rhombus, trapezoid, six-sided star, oval, and tear drop. Texture images were sourced from the Normalized Brodatz Texture Database (Abdelmounaime & Dong-Chen, 2013). These images included close-up photographs of various real-world textures, such as tree rings, sand dunes, snakeskin, bubbles, etc. Subjects did not report difficulty in discriminating between textures (Figure 2.1A). The stimuli generally subtended ~7.5˚ of visual angle. Stimulus position in the hierarchical blocks was computed along an invisible circle positioned at the center of the screen with a radius subtending ~7.5˚ of visual angle. The eight locations along this circle began at 27.5˚ clockwise from the vertical meridian and were equally spaced by 45˚ increments.

Flat Block

The flat block consisted of 20 repetitions of each stimulus for a total of 160 trials. Stimulus order was randomized within each set of 8 trials so as to restrict the range of the number of trials between stimulus repetitions. On average, each stimulus was viewed once every 8 trials, ranging from 0 to 15. Prior to the start of the block, subjects had the opportunity to view all 8 stimuli created for the upcoming block. All stimuli were presented on screen in a 2x4 array and remained on screen until the subject chose to proceed. No additional instructions were provided regarding the viewing of the stimuli.

Trials began with the presentation of the stimulus slightly offset left of the center of the screen for a maximum of 2,000ms (Figure 2.1B). Stimulus composition included a black-and-white image cropped into a specific shape with a colored border. Subjects were instructed to respond to the presentation of the stimulus by pressing one of four buttons mapped to their right index, middle, ring, and pinky fingers. Responding within 2,000ms advanced the trial to the confidence response phase. This phase began with the appearance of a vertical rectangle offset right of center with a horizontal black bar appearing either on the bottom or top of the rectangle. To indicate their confidence that their most recent response to the stimulus was correct, subjects had 1,500ms to re-press and hold down the button they had just pressed. By re-pressing the button, the black bar began to move away from its starting position at a constant rate until it reached the other side of the rectangle, a process that lasted up to 1,250ms. Regardless of the bar's starting position, the top of the rectangle indicated 100% confidence in their answer being correct, while the bottom of the rectangle indicated 0%. Subjects were instructed to be as precise as possible with their confidence rating. Following the release of the held-down button, or after 1,250ms, both the stimulus and confidence probe disappeared from screen. Following a pseudo-random inter-stimulus interval (200ms, 1,200ms, or 2,200ms), subjects received audiovisual feedback. Correct feedback involved the presentation of the word "Correct" and "$$$$$" stacked vertically in the center of the screen, as well as a pleasant tone. Incorrect feedback contained the word "Incorrect" and an unpleasant tone. The feedback stimulus persisted for 333ms. Feedback

was 100% valid. Following feedback, a fixation cross was displayed for the remainder of the trial duration (5,283ms, 6,283ms, or 7,283ms, depending on the duration of the inter-stimulus interval on that trial). The next trial then began after a variable inter-trial interval (ITI) with a mean of 1,500ms (range = 500ms – 4,500ms). The order of ITIs within a block was optimized to permit estimation of the event-related response using optseq2 (Dale, 1999).

We assigned two stimuli to each response option so that each button had a 25% chance of being correct on any given trial. Stimulus-response mappings were independent from one another, such that no higher-order structure was present, thus requiring each response to be learned individually. Following the final trial, mean block accuracy was presented on screen.

Hierarchical Block

We designed the hierarchical blocks identically to the flat block with the following exceptions. (1) Stimuli now included a fourth dimension: position on screen. In each of the four hierarchical blocks, the stimulus could appear in one of two locations on screen. These locations were semi-randomly selected from 8 possible equidistant positions along an invisible aperture around the center of the screen. We assigned the positions in each block in pairs, such that each pair was offset in both the x- and y-axis so as to create as large a separation and difference as possible. Position was not included in the flat block as pilot testing indicated subjects were unable to learn above chance 16 independent stimuli across four button responses in an appropriate amount of time. (2) The number of stimulus repetitions decreased from 20 to 6, resulting in a decrease in the number of total trials from 160 to 96 per block. (3) Given the new position dimension, the confidence probe was moved to the center of the screen so as not to interfere with the stimulus. (4) The position-on-screen dimension was not included in the pre-block stimulus presentation screen in which all 8 stimuli were shown.

Lastly, and most critically, all hierarchical blocks contained a 2nd-order policy relationship that subjects could discover and transfer across blocks so as to facilitate their learning, instead of learning 16 independent stimulus-response mappings. Specifically, the shape dimension cued 1st-order rules dependent on either the colors or textures, and as a result, screen position was irrelevant. By learning and exploiting this structure the number of rules to be learned decreased to four (i.e. two rules for color, two rules for texture). The same 2nd-order policy relationship was maintained across blocks, in that the shape dimension (shape) always cued rules based on either color or texture dimensions.

Instructions and Training Protocol

Prior to performing the task inside the MRI scanner, all subjects completed a training session on a desktop computer to make sure they understood the task and could perform it adequately. After obtaining experimental consent, and confirming both study and MRI scanner eligibility, subjects reviewed the instructions of the task. Along with visual aids on the computer, the experimenter described the task such that subjects knew they had to learn stimulus-response mappings across multiple task blocks, however no information was provided that could cue subjects to the hierarchical structure of the task. Subjects then practiced the confidence-reporting component of the task in a guided environment using stimuli not present in the real experiment.

Subjects received guided instructions indicating which button to press and how confident they should report feeling for each practice trial. Instructed confidence levels included 0%, 15%, 35%, 50%, 65%, 85%, and 100%. Subjects needed to place the confidence bar at the appropriate location along the vertical rectangle to match the instructed confidence level across 21 practice trials (3 repetitions of each level). A 93% accuracy criterion was required to progress. Subjects had to repeat the 21-trial practice block until they met criterion. The timing of all events matched that of the real experiment.

Following completion of the confidence reporting practice, subjects then performed 24 practice trials of a flat block, using the same stimuli as before. Just as in the real task, subjects had to learn eight independent stimulus response mappings across four buttons using the feedback provided at the end of each trial. No performance criterion was included, as the goal of this practice session was to familiarize subjects with the components of the task in real time.

Upon completion of the practice session, subjects were then escorted to the MRI scanner suite and placed inside the scanner. During the acquisition of an anatomical scan (details below), subjects went through the practice instructions and confidence reporting session again so as to become accustomed to both the MRI-compatible four-button response box, and to being inside the active scanner. Subjects received compensation at a rate of $20 per hour and could earn a bonus of up to $10 based on their overall trial accuracy.

Statistical Analyses of Behavioral Data

Analyses of behavioral data included the use of paired t-tests with one exception. When analyzing the number of learned 2nd-order rules across blocks, we used Wilcoxon sign-ranked tests due to the non-parametric nature of the data (i.e. subjects could learn either zero, one, or two 2nd-order rules per block) and the within-subjects design of the study. In addition, the stimulus dimension of position-on-screen was fully ignored in all analyses of the data.

Statistical Analyses of fMRI Data

Whole-brain analyses were performed in SPM and cluster correction was performed at the family wise error rate of $p = 0.05$, using $p = 0.001$ as the cluster defining threshold. Correlations between fMRI data and behavioral data were performed using standard parametric linear regression, as well as non-parametric rank-ordered regression in order to better control for potential outliers in the dataset. Results for each assessment are presented in tandem throughout the manuscript.

fMRI Data Acquisition

Whole-brain imaging was performed at the Henry H. Wheeler Jr. Brain Imaging Center at UC Berkeley using a Seimens 3T Trio MRI scanner using a 32-channel head coil. Functional imaging data was acquired with a gradient-echo echo-planar pulse sequence using a multi-band acceleration factor of 4 (TR = 1,000ms, TE = 33ms, flip angle = 40˚, array = 84 x 84, 52 slices, voxel size = 2.5mm isotropic). T1-weighted MP-RAGE anatomical images were collected as well (TR = 2,300ms, TE = 2.98ms, flip angle = 9˚, array = 256 x 256, 160 slices, voxel size =

1mm isotropic). Subject's head movement was restricted using foam padding. Auditory feedback was presented through in-ear headphones connected to the stimulus presentation computer. The flat block consisted of a single run of 1290 TRs, while each hierarchical block consisted of 760 TRs.

fMRI Data Preprocessing

Preprocessing was performed using FMRIPREP v1.0.2 (Esteban et al., 2018), a Nipype (Gorgolewski et al., 2011) based tool. Each T1w (T1-weighted) volume was corrected for INU (intensity non-uniformity) using N4BiasFieldCorrection v2.1.0 (Tustison et al., 2010) and skull-stripped using ANTs BrainExtraction. Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c was performed through nonlinear registration with the antsRegistration tool of ANTs v2.1.0 (Avants, Epstein, Grossman, & Gee, 2008), using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using FSL's fast (Y. Zhang, Brady, & Smith, 2001). Functional data were motion corrected using FSL's mcflirt (Jenkinson, Bannister, Brady, & Smith, 2002b). This was followed by co-registration to the corresponding T1w using boundary-based registration (Greve & Fischl, 2009a) with 9 degrees of freedom, using flirt (FSL). Motion correcting transformations, BOLD-to-T1w transformation, and T1w-to-template (MNI) warp were concatenated and applied in a single step using ANTs ApplyTransforms using Lanczos interpolation. Slice timing correction was not performed. Preprocessed data were spatially smoothed with an 8mm FWHM isotropic Gaussian kernel. Motion estimates used for subject exclusion were calculated using SPM's realign function.

Computational Modeling: State-Space Model

Trial responses were modeled with a state-space modeling approach (A. C. Smith et al., 2004) to produce learning curves. The model outputs trial-by-trial estimates of the probability of a correct response on each trial, as well as a 90% confidence interval around each estimate. Similar to (Badre et al., 2010), our analyses focused on the following metrics derived from the learning curve: (1) the trial for which the 90% confidence interval no longer included chance performance, referred to as the "learning trial"; (2) the maximal 1st derivative of the learning curve, which indexes the rate of learning; (3) the maximal 2nd derivative, which indexes the rate of change in one's learning rate.

Computational Modeling: Mixture of Experts Model

We make use of a hybrid Bayesian-reinforcement learning mixture of experts (MoE) model previously used by Frank and Badre (2012) in order to estimate subjects' attention to various hypothesis states that we assume are being tested while subjects perform the task. Given the observed stimuli and responses, the MoE model estimates individual subjects' attention to likely hypotheses about the relationship between context (i.e. the features of the stimulus) and action (i.e. the available button responses) in each task block. Each expert in the model represents a prediction about how a stimulus feature, or combination of features, relates to the likelihood of obtaining a reward given the motor actions available to the subject. For example,

the "shape expert" could learn the likelihood of obtaining a reward based only on the shape of the stimulus. For each trial, the expert makes its prediction about what action is likely to be correct given its assigned feature, and experts who contribute accurate predictions are rewarded while experts providing unreliable predictions are not. For hierarchical experts, the model makes predictions about subordinate stimulus dimensions (i.e. color or texture) contingent on the identity of a third, superordinate dimension (i.e. shape), such that weights assigned to predictions about each subordinate dimension are dynamically gated based on the feature of the superordinate dimension (e.g., circle vs. square). The MoE model also assigns attentional weights to experts that learn the overall reliability of hierarchical vs. flat predictions based on the reliability of all the hierarchical and flat experts, respectively.

For the current study, we adapted the model in order to allow for individual fits to each hierarchical expert. As the original version used a single hyperparameter across all three hierarchical experts, thus preventing the ability to estimate different initial weights, we instead modeled each hierarchical expert with a separate parameter. We also removed the decay parameter originally used to model the degree to which the current block's attentional weights carried over into the next block. Instead, we modeled a separate set of parameters for the various experts in each block. By removing the decay parameter, and modeling each block independently, we ensure that the model is incapable of being biased by the previous block. As a result, any differences between blocks in the parameter values, as well as the computed attentional weights at the beginning of the block, are the result of that block's data alone.

Specifically, subjects' beliefs about reward probability for each of the 4 available response options (per expert) were modeled as a Beta distribution and updated via Bayes' Rule. For example, the color expert was updated by:

$$P\left(\theta_{R,C}\middle|r_1 \dots r_n\right) \propto P(r_1 \dots r_n\middle|\theta_{R,C})P(\theta_{R,C})$$

where $\theta_{R,C}$ reflects the parameters determining the belief distribution about rewards given the presence of color C and the choice of response R, with $r_1 \dots r_n$ being the rewards seen so far when this specific R was chosen. Next, the probability of selecting each response is calculated by comparing the means μ of their reward distributions using a softmax function. For example, the probability of the color expert selecting R_i on trial t was

$$P_{R_i}^{C}(t) = \frac{e^{\frac{\mu_{Ri}^{C}(t)}{\kappa}}}{\sum_j e^{\frac{\mu_{Rj}^{C}(t)}{\kappa}}}$$

where κ governs the choice stochasticity, with lower (higher) values reflecting less (more) noise. The same computations were performed for each expert e, including a shape expert and texture expert, as well as all 2-way conjunctions, and finally the full 3-way conjunction. The model represents subjects' beliefs about the reliability of each expert with another Beta distribution, and again uses Bayes' Rule to learn the probability that the expert is reliable. For example, the color expert is updated as follows

$$P(\theta_C|r_1' \dots r_n') \propto P(r_1' \dots r_n'|\theta_C)P(\theta_C)$$

where *r'* are the rewards indicating whether the expert contributed to the outcome. Specifically, if R_i is the chosen action, rewards are delivered as follows

$$r = \begin{cases} r, & if\ \mu_{R_i} > \mu_{R_j}, \forall j \neq i \\ 1 - r, & otherwise \end{cases}.$$

Thus, experts were rewarded when a reward was received and that expert assigned the highest probability to the chosen response. If, on the other hand, the expert predicted one of the unselected options it would not be rewarded (i.e. *r* = 0). Moreover, if the chosen action was not correct and the expert assigned the largest probability to that action, then it was not rewarded. However, it was rewarded if the outcome was not correct (i.e. it did not contribute to the incorrect action). We can assign an attentional weight to each expert that reflects its history of contributing to successful outcomes. To do so, we use another softmax function to assign weights to each expert, relative to all other experts. For the color expert, we can determine its weight with the following equation:

$$w_C(t) = \frac{e^{\frac{\mu^C(t)}{\zeta}}}{\sum_E e^{\frac{\mu^E(t)}{\zeta}}}$$

where w_C on trial *t* is the attentional weight, based on its expected reward probability μ^C relative to all other experts. Lastly, ζ acts as a gain parameter that discriminates between the separate experts (similar to the κ parameter in the action selection softmax). Thus, the probability of selecting response R_i is the sum of the experts *E* in proportion to their weight

$$P_{R_i}^f(t) = \sum_E w_E P_{R_i}^E(t)$$

where P^f refers to the probability of generating responses for a superordinate "flat expert" (the combination of the all subordinate experts so far mentioned).

At this point, the model is incapable of detecting any hierarchical structure that may be present in the task. In order to afford the model this ability, we now discuss the inclusion of a set of "hierarchical experts". These experts learn about 2 of the stimulus dimensions contingent on the identity of another, higher order dimension. For example, the texture hierarchical $h_{CS|T}$ expert would learn reward probabilities for color and shape separately for each texture option in *T*. This manner of learning is accomplished by having 2 subordinate experts learn the reward probability for selecting a response for color *C* (shape *S*) given texture *T*:

$$P(\theta_{R,C|T}|r_1 \dots r_n) \propto P(r_1 \dots r_n|\theta_{R,C|T})P(\theta_{R,C|T})$$
$$P(\theta_{R,S|T}|r_1 \dots r_n) \propto P(r_1 \dots r_n|\theta_{R,S|T})P(\theta_{R,S|T}).$$

Credit assignment works as it did with the flat experts, but now across the subordinate experts within the hierarchical expert framework. For the $h_{CS|T}$ hierarchical texture expert, attentional weights are dynamically assigned to the color or shape dependent on the texture:

$$w_{C|T}(t) = \frac{e^{\frac{\mu^{C|T}(t)}{\zeta}}}{e^{\frac{\mu^{C|T}(t)}{\zeta}} + e^{\frac{\mu^{S|T}(t)}{\zeta}}}$$

where $w_{C|T}(t)$ is the attentional weight to the color expert relative to the shape expert when texture *T* is present. The probability of selecting a response R_i according to this hierarchical texture expert is the result of mixing the subordinate experts on each trial:

$$P_{R_i}^{h_{CS|T}}(t) = w_{C|T}\, P_{R_i}^{C|T}(t) + w_{S|T}\, P_{R_i}^{S|T}(t).$$

In addition to the texture expert, we also included a hierarchical shape and hierarchical color expert. Similar to the overall flat expert, a superordinate hierarchical expert assigned attention weights to the hierarchical experts via:

$$P_{R_i}^{h}(t) = w_{CS|T}\, P_{R_i}^{CS|T}(t) + w_{CT|S}\, P_{R_i}^{CT|S}(t) + w_{TS|C}\, P_{R_i}^{TS|C}(t).$$

Lastly, a second-level attentional selection step was included to arbitrate between the two overall experts (flat, hierarchical):

$$w_H(t) = \frac{e^{\frac{\mu^{H}(t)}{\xi}}}{e^{\frac{\mu^{H}(t)}{\xi}} + e^{\frac{\mu^{F}(t)}{\xi}}}$$

where ξ determines the gain of the discrimination between the hierarchical and flat expert. The ultimate response is then selected as follows:

$$P_{R_i}(t) = w_H\, P_{R_i}^{H}(t) + w_F\, P_{R_i}^{F}(t).$$

In total, the model included 11 free parameters to be estimated, with each block being fit independently. Three of these consisted of the α-parameters from each one-way flat experts' initial Beta distribution (the mean of which is represented by, in the example of the flat color expert, μ^C). Another two came from the β-parameter of the Beta distribution for the two-way and three-way flat experts, where in the case of the three two-way experts, the value acted as a hyperparameter over each expert. Another 3 consisted of the β-parameter of the Beta distribution for each hierarchical expert. The last three included the noise/gain parameters in each of the three softmax functions (i.e. action selection, attentional weight assignment, and superordinate attention to hierarchy).

In order to obtain the best fit for the data, we first modeled all subjects together (pseudo-R^2 = 0.25 and 0.12 for the mean hierarchical block and flat block, respectively) in order to

generate appropriate initial starting parameter values to be used as our initialization point for the model when fitting each subject individually (mean pseudo-R^2 = 0.33 and 0.15 for the mean hierarchical block and flat block, respectively). Model fitting occurred via maximum likelihood estimation. These pseudo-R^2 values are similar to those reported in Frank and Badre (2012). Validation of the revised MoE model involved simulating datasets across each of the five task blocks. We used the parameter values obtained from fitting the model to the real subject data to generate simulated responses to the task. In order to draw comparisons to the human data, the simulated data was then fit to the State Space model so as to produce learning curves, which allowed for calculation of learning metrics (i.e., maximum 2nd derivative). Overall, the revised MoE model was successfully able to recreate the qualitative patterns of behavior and attentional weight recovery across blocks seen in the human data.

Univariate fMRI Analysis

Statistical models were constructed for each subject under the assumptions of the general linear model (GLM) using SPM 12 (Statistical Parametric Mapping; www.fil.ion.ucl.ac.uk/spm). Each trial was modeled by one of two sets of five boxcar regressors: (1) a regressor for the stimulus response phase (beginning with stimulus onset and ending when a response was made), (2) a regressor with the same onset and duration as the stimulus response phase, but whose value was parametrically modulated by the subject's reaction time to the stimulus, (3) a regressor for the confidence response phase (beginning and ending with the onset and offset, respectively, of the confidence probe), (4) a regressor with the same onset and duration as the confidence response phase, but whose value was parametrically modulated by the reported confidence level, (5) a regressor for the feedback phase (beginning and ending with the onset and offset, respectively, of the audiovisual feedback). In order to match the analysis approach of Badre et al., 2010, one set of regressors exclusively modeled correct trials, while the other set exclusively modeled incorrect trials. To ensure the parametrically modulated regressors only explained the variance unique to processes associated with the modulatory values (i.e. stimulus reaction time and confidence level), we orthogonalized both the modulated stimulus response phase and modulated confidence response phase regressors with respect to their respective unmodulated regressors. Next, we included three additional regressors to remove variance associated with events related to the subject failing to make a required response. Two regressors modeled stimulus and confidence response phases where no stimulus or confidence response, respectively, was made. The third regressor modeled feedback phases where "No Response" was presented. Although trials where subjects failed to indicate their level of confidence could be separated by whether the subject's stimulus response was correct or incorrect, we chose to model these events together because we considered both events to be of no interest and thus nuisance signals. Lastly, five block regressors were included to account for run-to-run variance. In total, each block contained a theoretical maximum of fourteen regressors: some subjects had blocks where all required responses were made, and thus no regressors could be made that modeled events related to a failure to respond. Low frequency signals were removed with a 1/128 Hz high-pass filter. This first level regression thus yielded standardized regression coefficients ("betas") for each voxel in the brain for each regressor included in the model. Linear contrasts were used to obtain subject-specific effects, which were then entered into a second-level analysis treating subjects as a random effect and comparing voxel effects against a value of zero. Cluster correction was performed on all whole-brain, voxelwise analyses using an initial height threshold

of $p < 0.001$ in order to then define a familywise error rate threshold of $p = 0.05$. The first voxelwise analysis of stimulus response phase activity compared to baseline (Figure 2.3) resulted in an extent threshold of 29516 voxel. The voxelwise map revealing the contrast of stimulus response phase activity in Hier 2 greater than the average of Hier 1 and Hier 3 resulted in an extent threshold of 106 voxels (Figure 2.4A), while the voxelwise map assessing the behavioral metric of transfer in Hier 3 and Hier 4 resulted in an extent threshold of 107 voxels (Figure 2.4B).

Region-of-Interest (ROI) analyses supplemented the whole-brain search. ROIs were constructed with the Marsbars (Brett, Anton, Valabregue, & Poline, 2002) and wfupickatlas (Maldjian, Laurienti, Kraft, & Burdette, 2003) toolboxes in SPM12. Coordinates and sphere size for frontal cortex nodes (i.e. dorsal premotor cortex (PMd), pre-dorsal premotor cortex (pre-PMd), mid inferior frontal sulcus (Mid-IFS), and frontal polar cortex (FPC)) were taken from Badre et al., (2010). Cingulo-opercular and fronto-parietal coordinates and size (i.e. CO: bilateral anterior prefrontal cortex, bilateral anterior insula / frontal operculum, bilateral thalamus, and dorsal anterior cingulate cortex / mid-superior frontal cortex; FP: bilateral intraparietal sulcus, bilateral frontal cortex (roughly BA 6), bilateral precuneus, bilateral inferior parietal lobule, bilateral dorsolateral prefrontal cortex (roughly BA 9/46), and midcingulate cortex) were taken from Dosenbach et al., (2007).

Behavioral Metrics of Transfer

To test for brain-behavior correlations that relate individual differences in transfer performance to fMRI activity, we calculated the behavioral metric of transfer based on the state-space model we employed. We computed a difference score between the fourth and the first hierarchical block so as to assess the maximum impact that hierarchical structure transfer could have on behavioral performance. Specifically, our metric of transfer came from computing the change in the state-space model's maximum 2^{nd} derivative measure. We chose to focus on the maximum 2^{nd} derivative as it should best capture the degree to which learning accelerates once the subject determines the appropriate 1^{st}-order rules associated with the known 2^{nd}-order policy. Defining transfer in this manner allowed us to contrast subjects' performance when learning a hierarchically structured task with no ability to transfer knowledge of 2^{nd}-order policy to when subjects have the greatest likelihood of transferring learned 2^{nd}-order policy.

For the whole-brain analysis, we defined a contrast for each subject that contrasted mean stimulus response phase activity for the third and fourth hierarchical block against baseline. At the second level, the transfer metric was used as a covariate and regressed against this contrast to identify univariate activity across individuals that was associated with differences in transfer.

2.4 Results

State-Space Model Reveals Discovery and Transfer of Global Hierarchical Structure

Trial outcomes from each block were fit with a state-space model (Figure 2.2A, Smith et al., 2004), and the following metrics were computed from the learning curves in each block: the (1) maximal 1st derivative (Figure 2.2B), (2) maximal 2nd derivative (Figure 2.2B), and (3) the "learning trial" (Figure 2.2B) (see Material and Methods for definitions).

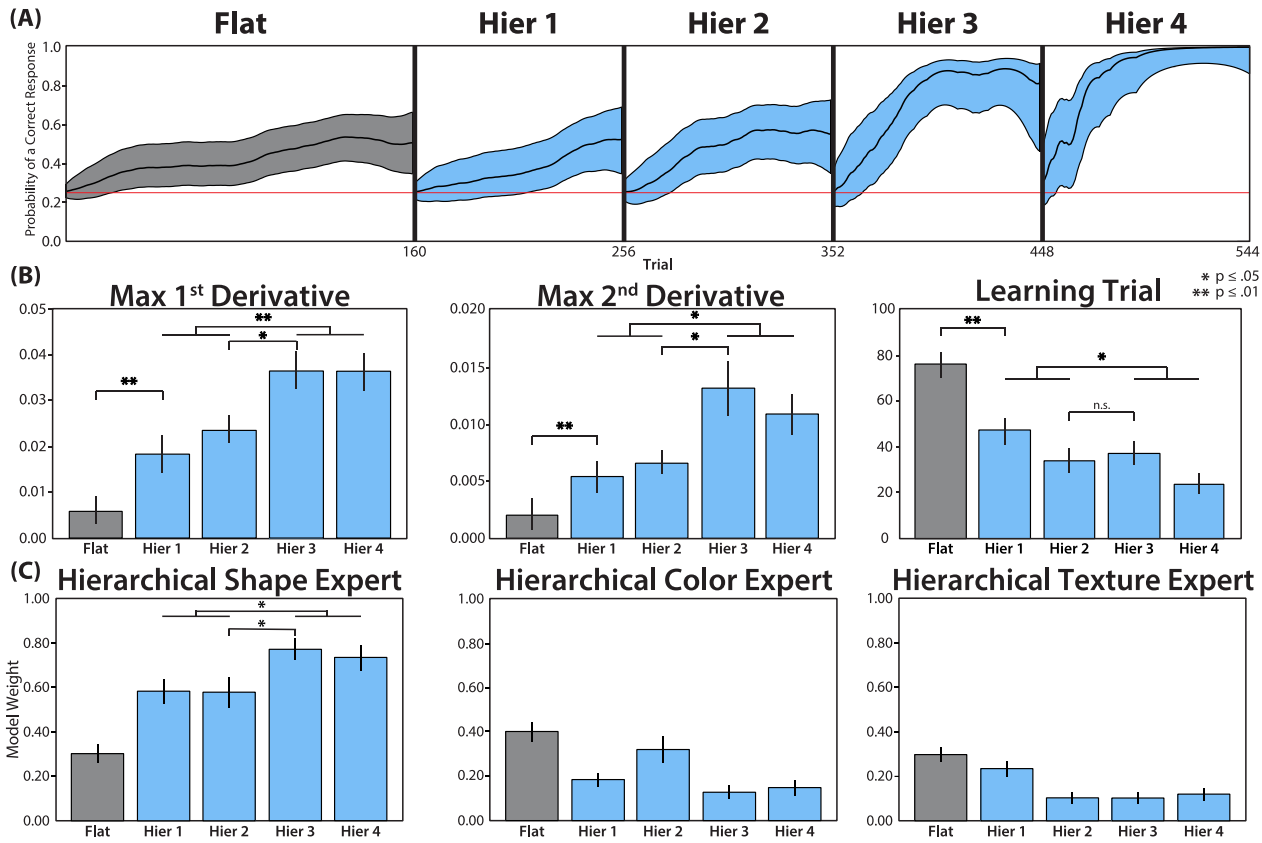


Figure 2.2 Learning Curve and Mixture of Experts Results Reveal the Discovery and Transfer of the Global Hierarchical Policy Structure. (A) Output of the state-space model (Smith et al., 2004) for a representative subject. For each trial within a block, the model computes the probability of a correct response given the block's trial outcomes. The 90% confidence interval around each trial's estimated probability is shown in grey (Flat block) and blue (Hierarchical blocks). The red line indicates chance-level performance. (B) State Space model estimates for maximal 1st and 2nd derivatives and learning trial, averaged across subjects. The 1st and 2nd derivative metrics reveal a significant increase in learning following the 2nd hierarchical block, while the mean learning trial improves more gradually across hierarchical blocks. (C) Mixture of Experts model weights for Attention to the Hierarchical Shape, Color, and Texture Experts at the beginning of the Flat (gray) and four Hierarchical (blue) blocks. Each expert corresponds to a latent hypothesis regarding the hierarchical task structure that a subject might hold at the beginning of each block. Following the second hierarchical block, there is a significant increase in attention for the expert that corresponds to the global 2nd-order policy: shape cues color or texture (Hierarchical Shape Expert). Error bars represent within-subjects standard error of the mean. Significance is assessed at $p < 0.05$.

We first tested whether subjects acquired 2nd-order hierarchical rules in blocks that contained a hierarchical policy structure, which should be reflected in differences in the learning curve metrics. Compared to the flat block, learning in the first hierarchical block was more efficient (earlier learning trial: $t(23) = 3.22$, $p = .004$; Figure 2.2B) and showed the abrupt gains in accuracy expected from generalization of learned 2nd-order policy to unknown 1st-order rules (greater max 1st derivative: $t(23) = 3.30$, $p = 0.003$; max 2nd derivative: $t(23) = 3.20$, $p = 0.004$; Figure 2.2B). This pattern was also present when comparing learning curve metrics from the flat

block to the average metrics across all hierarchical blocks: max 1st derivative: $t(23) = 5.74$, $p < 0.001$; max 2nd derivative: $t(23) = 4.68$, $p < 0.001$; learning trial: $t(23) = 3.95$, $p < 0.001$, Figure 2.2B).

We next sought to investigate the role of hierarchical structure transfer. In the first hierarchical block, subjects acquired and exploited the block-specific 2nd-order policy to facilitate learning relative to the flat block. Subsequently, the second hierarchical block provides the opportunity for subjects to discover the global 2nd-order policy structure: after acquiring the block-specific 2nd-order policy in the second hierarchical block, subjects can discover that the same abstract 2nd-order policy (i.e., shape cues color or texture) has been shared across the first two hierarchical blocks. Subjects can then transfer their learned knowledge of a global 2nd-order policy structure to subsequent blocks, which should greatly facilitate the acquisition of a block-specific 2nd-order policy (e.g. star cues color, trapezoid cues texture) and subsequently allow the subject to more rapidly resolve 1st-order rules within the known hierarchical structure. Thus, we predicted that successful structure transfer would result in markedly more efficient and abrupt learning following the second hierarchical block.

To test for behavioral evidence of hierarchical structure transfer, performance in hierarchical block three – where subjects can implement learned structure knowledge from the start of the block – was compared to hierarchical block two – where subjects can initially discover the global 2nd-order policy structure (Figure 2.2B). As predicted, there is a significant improvement in hierarchical learning as measured by the max 1st derivative ($t(23) = 2.25$, $p = 0.035$), and max 2nd derivative ($t(23) = 2.23$, $p = 0.036$). However, the learning trial metric does not show the same pattern ($t(23) = 0.41$, $p = 0.688$). This improvement is not easily explained by general practice effects: there is not a reliable change in performance metrics from the first to the second hierarchical block – when subjects can take advantage of task practice and general familiarity with the trial procedure – but must still discover the global 2nd-order policy structure (as assessed by all three metrics, max $t = 1.28$, $p = 0.21$). Instead, the evidence of transfer is only observed after subjects have had the opportunity to discover the global structure in the second hierarchy block.

Following discovery of the global 2nd-order policy structure, hierarchical knowledge transfer can facilitate learning for all subsequent blocks. Therefore, the learning metrics averaged across hierarchical blocks three and four (when the knowledge can be implemented to support learning) were compared to the average across hierarchical blocks one and two (when the knowledge has not yet been acquired). Subjects showed evidence of improved hierarchical learning in the last two hierarchical blocks versus the first two across all behavioral metrics: max 1st derivative: $t(23) = 2.99$, $p = 0.007$; max 2nd derivative: $t(23) = 2.76$, $p = 0.011$; learning trial: $t(23) = 2.50$, $p = 0.020$.

Lastly, we used a method previously developed to assess hierarchical learning (Badre et al., 2010) to analyze hierarchical structure learning and transfer. Instead of modeling all trials together within a block, responses to each unique stimulus were individually analyzed in order to obtain separate learning trials. Moreover, in tasks with hierarchically structured 2nd-ordered policy, one can conclude that a 2nd-order rule is completely learned if all of its subordinate 1st-order rules are learned above chance. Then, evidence of hierarchical structure transfer can be

assessed, which should allow for faster and more complete learning of 2nd-order rules. Subjects learned more 2nd-order rules in the hierarchical blocks than in the flat block ($Z = 15.5$, $p < 0.001$). Moreover, there was a significant increase in learned 2nd-order rules from the second to the third hierarchical block ($Z = 4.5$, $p = 0.008$). Lastly, subjects also learned more 2nd-order rules in the last two hierarchical blocks than in the first two ($Z = 12.0$, $p < 0.001$). Together, these results provide evidence that learning and subsequently transferring the global 2nd-order policy structure supports more efficient hierarchical learning, over and above the expected level of hierarchical learning if the hierarchical policy must be re-learned on every block.

Mixture of Experts Model Confirms Transfer of Specific Hierarchical Structure

Although learning rate metrics derived from the state-space model allow us to characterize how learning changes across blocks, they do not provide information about *why* learning may have changed. We theorized that subjects discovered the specific 2nd-order policy that was globally persistent across blocks. When learning the rules for a new block, this knowledge should encourage subjects to test the hypothesis that shape determines 2nd-order policy. In turn, this would enhance learning by biasing their attention toward the relevance of the shape dimension, and away from the color and texture dimensions. As an alternative explanation, subjects might have discovered that the presence of hierarchical policy, in general, was persistent across blocks: one dimension cues the relevant 1st-order dimensions. When learning the rules for a new block, this knowledge should encourage subjects to test the hypothesis that a 2nd-order policy exists. This knowledge could enhance learning by biasing their attention towards the relevance of 2nd-order policies, in general, versus a flat policy. Because the state-space model cannot distinguish these two explanations, we used a hybrid Bayesian-reinforcement learning Mixture of Experts (MoE) model to infer the latent hypothesis states of each subject during the learning process (Frank & Badre, 2012). This approach allows us to probe the underlying cognitive mechanisms that support transfer by estimating how specific hypotheses regarding hierarchical task structure were being attended and transferred across blocks (see Materials and Methods for details).

The MoE model was employed to derive attention measures for four modeled "experts" each associated with a specific hypothesis. The first measure indexes the attention subjects place on the specific hypothesis that the shape dimension forms the top of the 2nd-order policy and cues subordinate 1st-order rules based on either color or texture (referred to as "attention to the hierarchical shape expert"). The second and third measures index the attention placed on the specific hypotheses that the color or texture dimensions, respectively, form the top of the hierarchy. The fourth measure indexes the attention subjects place on the general hypothesis that hierarchical structure, in the form of any 2nd-order policy, exists in the block compared to a flat policy (referred to as "attention to hierarchy"). The attention to hierarchy measure does not discern between which dimension sits atop the hierarchy, in contrast to the other three measures. In order to characterize what knowledge is being transferred from the previous block, we focus on the model estimates for these measures that capture the state of the subject prior to encountering the first trial of the block. These estimates of the subject's latent state before the block begins are inferred by fitting the model to each individual's trial-by-trial sequence of choices and rewards. Therefore, a discrimination can be made between whether a subject is transferring a hypothesis regarding a specific 2nd-order policy (attention to the hierarchical shape

expert), compared to a general hypothesis regarding the presence of 2nd-order policy (attention to hierarchy), at the start of the block.

First, in order to determine whether subjects discover the global 2nd-order policy that is persistent across blocks and then test the hypothesis that this policy applies to subsequent blocks, the attention to the hierarchical shape expert was analyzed across blocks (Figure 2.2C). In line with our predictions, subjects' attention to the hierarchical shape expert at the start of the block increases from the second to the third hierarchical block ($t(23) = 2.08$, $p = 0.049$, Figure 2.2C), after they have had the opportunity to discover the global 2nd-order policy structure. Moreover, because this knowledge can inform the hypotheses for all subsequent blocks, attention to the hierarchical shape expert is greater at the start of hierarchical blocks three and four than at the start of the first two hierarchical blocks ($t(23) = 2.64$, $p = 0.015$). Although specific statistical predictions regarding attention to the hierarchical color and texture experts were not made, attention to these experts should generally be diminished when attention is biased in favor of the hierarchical shape expert. Indeed, attention to the color and texture experts is qualitatively low in the hierarchical blocks (Figure 2.2C).

Next, we analyzed whether subjects test the hypothesis that a hierarchical policy, in general, is persistent across blocks. Subjects' attention to hierarchy does not increase from the second to the third hierarchical block ($t(23) = 1.70$, $p = 0.103$). However, there is a more gradual change in attention to hierarchy such that the measure increases from the first two hierarchical blocks to the last two ($t(23) = 2.51$, $p = 0.019$). Together, these results show that the improvement in hierarchical learning observed after the second hierarchical block can be explained by subjects discovering and then transferring their knowledge of the appropriate global 2nd-order policy structure that is persistent across all hierarchical blocks.

Lateral Frontal Regions Linked to Discovery of Global Hierarchical Structure

First, a whole-brain univariate contrast of activity during the stimulus response phase on correct trials across all blocks compared to baseline was performed ($p = 0.05$ cluster-corrected, Figure 2.3A). The resultant map is consistent with those seen in previous hierarchical reinforcement learning studies (Badre et al., 2010). The task recruited regions along the lateral frontal cortex associated with hierarchical task performance (Badre & D'Esposito, 2007; Koechlin, 2003), as well as parietal cortex, and more specifically the intraparietal sulcus, anterior insula, mid-cingulate cortex, occipital lobe, thalamus, and medial temporal lobe.

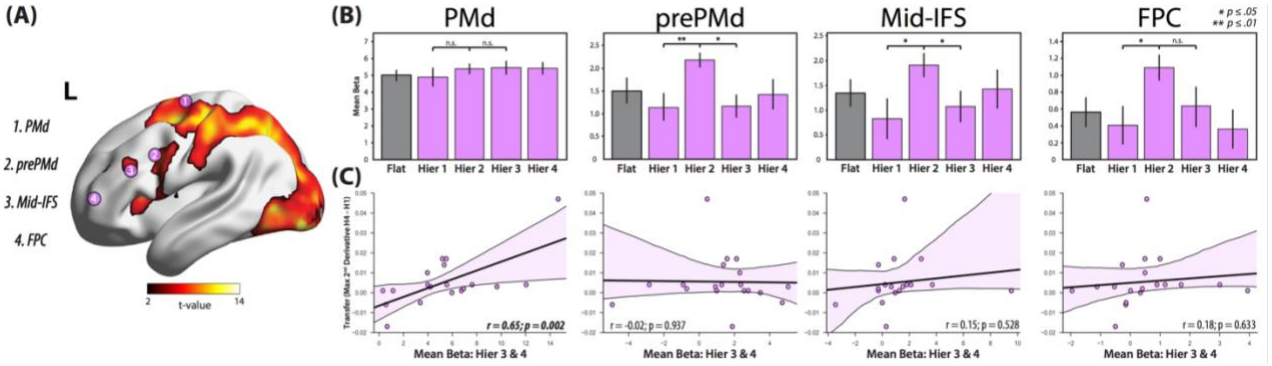


Figure 2.3 Lateral Frontal Regions Linked to Discovery of Global Hierarchical Policy Structure and Behavioral Transfer. (A) Group-level activity across all blocks during the stimulus response phase on correct trials only. The overlaid numbered pink circles indicate the position of each of the four lateral frontal cortex ROIs. Map is cluster-corrected to a family-wise error rate of $p < .05$. (B) ROI analyses for the regions shown in A. The mean beta coefficients from the stimulus response phase show elevated activity during the second hierarchical block (versus the first and third hierarchical blocks) in all regions except PMd. Error bars indicate within-subject standard error. (C) Correlations between behavioral transfer and activity in the left lateral frontal cortex ROIs following discovery of the global hierarchical policy structure. Only activity in PMd is tentatively correlated with individual differences in transfer.

To address which brain regions supported searching for and discovering the global 2nd-order policy structure, we first focused on regions in left lateral frontal cortex that support the learning and execution of hierarchical control policies: the dorsal premotor cortex (PMd), pre-dorsal premotor cortex (prePMd), mid-inferior frontal sulcus (Mid-IFS), and frontal polar cortex (FPC) (Badre et al., 2010, Figure 2.3A). In their original work, Badre and colleagues (2007) discovered that PMd resolved competition between 1st-order rules regarding motor response options, prePMd resolved competition between 2nd-order rules relating one stimulus feature to another (e.g. for squares, red cues action 1 while blue cues action 2), Mid-IFS resolved competition between 3rd-order rules, and FPC resolved competition between 4th-order task contexts. Moreover, activity in these regions has been associated with the search for a specific hierarchical policy within a task block (Badre et al., 2010). However, it remains unknown whether these same regions also support the learning of a more abstract, global hierarchical structure that facilities learning the specific hierarchical policies within each block.

The behavioral results demonstrate that subjects were able to both learn block-specific hierarchical policies, as well as search for and discover the global hierarchical policy structure during the second hierarchical block. To identify activity in the frontal cortex that is related to discovering the global structure, over and above activity associated with learning a block-specific hierarchical policy, activity in the second hierarchical block relative to the first hierarchical block was assessed (Figure 2.3B). With the exception of PMd ($t(18) = 0.72$, $p = 0.483$), activity across the lateral frontal cortex regions is greater in the second hierarchical block compared to the first (prePMd $t(18) = 3.48$, $p = 0.003$; Mid-IFS $t(18) = 2.10$, $p = 0.050$; FPC $t(18) = 2.19$, $p = 0.042$). Next, activity in the second hierarchical block was compared to the third hierarchical block, where subjects no longer need to search for structure and can instead implement their transferred structure knowledge from the second hierarchical block (Figure 2.3B). Again, activity in prePMd ($t(18) = 2.80$, $p = 0.012$) and Mid-IFS ($t(18) = 2.12$, $p = 0.048$) is greater in the second hierarchical block. Activity is also numerically greater in FPC ($t(18) = 1.52$, $p = 0.147$), but not statistically significant. Lastly, activity in PMd did not differ across the

blocks ($t(18) = 0.12$, $p = 0.904$). Because the activity in rostral regions of frontal cortex is elevated in the second hierarchical block relative to both the preceding and proceeding blocks, the observed results are likely due to a process that is preferentially engaged in the second hierarchical block, as opposed to a process that continuously evolves over time such as effects related to time on task or practice.

Lateral Frontal Regions Linked to Transfer of Global Hierarchical Structure

Next, we determined if activity in the lateral frontal ROIs predicts behavioral transfer, which was indexed by more abrupt hierarchical learning in the blocks that follow discovery of the global hierarchical structure (see Materials and Methods for definitions and details). Different lateral frontal cortex regions could support transfer of the global hierarchical policy structure. For example, prePMd could support transfer of 2nd-order policy by means of a more efficient resolution of competition between competing within-block 2nd-order rules. Alternatively, if transfer is an additional third level in the policy hierarchy (i.e. the task block contextualizes 2nd-order rules associated with the shape dimension), then Mid-IFS (e.g. the region associated with policy abstraction one level greater than that being transferred) could support transferring learned structure. Lastly, FPC activity could support transfer, as structure transfer may be a form of extended temporal contextualization, or episodic control, that biases task representations across multiple blocks. Knowledge of the shape dimension's position in the hierarchy may take the role of a schema and thus recruit FPC to support the accommodation and contextualization of new information within this framework.

To test these predictions, correlations between the mean activity in each lateral frontal ROI from blocks where behavioral transfer could occur (i.e. hierarchical blocks three and four), and the behavioral metric of transfer for each subject were performed (Figure 2.3C). Activity in prePMd ($r = – 0.02$, $p = 0.937$), Mid-IFS ($r = 0.15$, $p = 0.528$), and FPC ($r = 0.18$, $p = 0.633$) did not reliably correlate with behavioral transfer. However, activity in the most caudal frontal region, PMd, appeared to reliably correlate with behavioral transfer ($r = 0.65$, $p = 0.002$). To test the robustness of these individual differences results, we also performed the analyses using a non-parametric rank-ordered regression test. In line with the previous results, PMd was significantly correlated with behavioral transfer (spearman rho = 0.48, $p = 0.037$), while prePMd, Mid-IFS, and FPC were not statistically significant (absolute value of all spearman rho's < 0.43, all p's > 0.065). However, one high-leverage subject who showed substantial behavioral transfer also had the highest activity in PMd (Fig. 3C). When this subject is removed from the analysis, the positive correlation no longer reaches statistical significance (PMd: $r = 0.37$, $p = 0.12$ (rho = 0.39, $p = 0.11$); all other ROI p's > 0.38 (non-parametric p's > 0.09)). Thus, these results suggest that PMd is the most likely lateral frontal region to relate to transfer, although this relationship may be modest and awaits confirmation in future studies.

Whole-brain Analyses: Regions Linked to Discovery of Global Hierarchical Structure

In order to identify regions recruited by the search and discovery of the global hierarchical policy structure, a whole-brain voxelwise analysis was performed by contrasting activity in the second hierarchical block to the average of the first and third hierarchical blocks (Figure 2.4A). This contrast revealed activity that overlapped with the left prePMd and Mid-IFS

ROIs. However, activity was also found in medial superior frontal gyrus, the left inferior parietal lobule (IPL) and intraparietal sulcus (IPS), and the right IPL. The location of these lateral frontal and parietal regions overlap with a set of regions referred to as the "FP network" that have been previously implicated in cognitive control functions (Dosenbach et al., 2007, 2008).

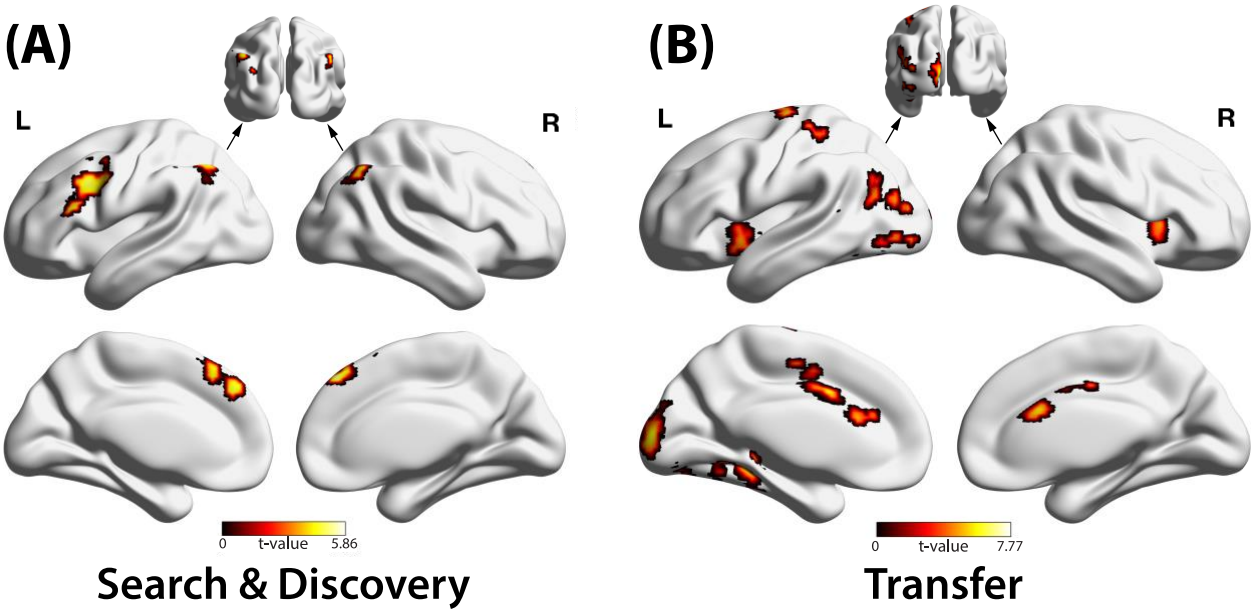


Figure 2.4 Voxelwise Analyses Reveal Regions Linked to Unique Behavioral Roles. (A) Activity during the search and discovery of the global hierarchical structure during the second hierarchical block shown by the contrast of Hier 2 > Hier 1 + Hier 3. (B) Whole-brain analysis of regions for which stimulus response phase activity following discovery of the global hierarchical structure correlates with behavioral transfer. All activity maps are cluster-corrected to a family-wise error rate of $p < .05$.

Whole-brain Analyses: Regions Linked to Transfer of Global Hierarchical Structure

To further identify which regions support behavioral transfer in blocks following the discovery of the global hierarchical policy structure, a whole-brain analysis was performed using the degree of behavioral transfer as a parametric modulator of the mean stimulus response phase activity in the third and fourth hierarchical blocks ($p = 0.05$ cluster-corrected, Figure 2.4B). PMd activity (overlapping with our ROI) – in accord with the previous ROI analyses – as well as bilateral anterior insula / frontal operculum, anterior cingulate cortex, left lateral occipital cortex, and left medial temporal cortex correlated with behavioral transfer. Anterior insula and dorsal anterior cingulate cortex correspond to the "core" regions of the putative cingulo-opercular network commonly found in tasks requiring cognitive control ("CO network", Dosenbach et al., 2007, 2008; Sadaghiani & Kleinschmidt, 2016).

Dissociation of Behavioral Roles for FP and CO Networks

The FP and CO networks have been proposed as two components of a dual-network architecture of cognitive control (Dosenbach et al., 2008), and regions in both the FP and CO networks were active during performance of our hierarchical learning task. However, these regions may support task performance by making separable behavioral contributions. To test this hypothesis, we directly compared the relationship between activity across the networks'

respective regions and (1) discovering the global hierarchical policy structure versus (2) transferring of hierarchical structure knowledge across blocks.

First, we assessed the relationship between activity in these networks and the search and discovery of hierarchical structure that occurs during the second hierarchical block. The canonical FP and CO networks were defined based on a previous meta-analysis of cognitive control tasks (FP: bilateral frontal cortex, bilateral dorsolateral prefrontal cortex, bilateral intraparietal sulcus, bilateral inferior parietal lobule, bilateral precuneus, and midcingulate cortex; CO: bilateral anterior insula / frontal operculum, bilateral anterior prefrontal cortex, bilateral thalamus, and dorsal anterior cingulate cortex / mid-superior frontal cortex, Figure 2.5A; coordinates from Dosenbach et al., 2007). Separately for the FP network and CO network ROIs, the activity during each block was estimated and a contrast was performed for the activity in the second hierarchical block versus the first and third hierarchical blocks (Figure 2.5B). FP activity was significantly increased during the second hierarchical block ($t(18) = 3.49$, $p = 0.003$), as expected based on the whole-brain results, whereas CO activity was not significantly different ($t(18) = 1.46$, $p = 0.162$). Since the FP network was chosen for further analysis based on the observation of left lateral frontal and bilateral parietal activity in our previous whole-brain contrast, any ROI analyses that include these regions may be biased by circularity (Vul, Harris, Winkielman, & Pashler, 2009). To address this possibility, a separate analysis was performed that included only the FP network ROIs that were not observed in the original whole-brain results (i.e. right frontal cortex, right dorsolateral frontal cortex, right intraparietal sulcus, bilateral precuneus, and midcingulate cortex), which also found a significant result for the contrast ($t(18) = 2.80$, $p = 0.012$). Next, to formally dissociate the patterns observed across the FP and CO networks (Henson, 2006), we tested the interaction between block (second hierarchical block; average of first and third hierarchical blocks) and region (FP; CO) and found that the difference in activity between the second hierarchical block compared to the first and third blocks is significantly greater in the FP regions than in the CO regions ($t = 3.37$, $p = 0.003$).

We next assessed the relationship between activity in these networks and the transfer of hierarchical structure knowledge. As before, we sought to confirm the relationship between behavioral transfer and the canonically defined CO network, while additionally ruling out the potential for circularity in our analyses. Our first analysis confirmed a significant relationship between activity averaged across all cingulo-opercular regions and behavioral transfer ($r = 0.57$, $p = 0.011$, spearman rho = 0.67, $p = 0.002$, Figure 2.5C). Moreover, in order to control for circularity in this analysis, we ran a separate test of the relationship between the CO network and behavioral transfer by excluding the insular and anterior cingulate ROIs that were present in the original whole-brain regression. This new analysis, which only included activity from bilateral thalamus and bilateral anterior prefrontal cortex (referred to as the "periphery" of the CO network, Dosenbach et al., 2008), found a significant correlation between mean ROI activity and the behavioral transfer metric ($r = 0.66$, $p = 0.002$, spearman rho = 0.77, $p < 0.001$). In contrast to the robust correlation between the CO network and behavioral transfer, activity in the FP network in the last two hierarchical blocks is only modestly correlated with the behavioral transfer metric ($r = 0.26$, $p = 0.279$, spearman rho = 0.61, $p = 0.006$, Figure 2.5C).

To test whether the CO network is uniquely related to transfer, both the CO network and FP network activity were included in a multiple regression with behavioral transfer as the

dependent variable, as this approach controls for any shared contribution made by both networks. This analysis revealed that activity in the CO network selectively predicts transfer (CO network: r = 0.52, p = 0.023, spearman rho = 0.47, p = 0.045; FP network: r = -0.003, p = 0.990, spearman rho = 0.04, p = 0.881, Figure 2.5D). Collectively, these findings demonstrate a clear dissociation: the regions of the FP network are specifically involved in the search and discovery of hierarchical structure, whereas the regions of the CO network are selectively involved in the transfer of hierarchical structure knowledge across blocks.

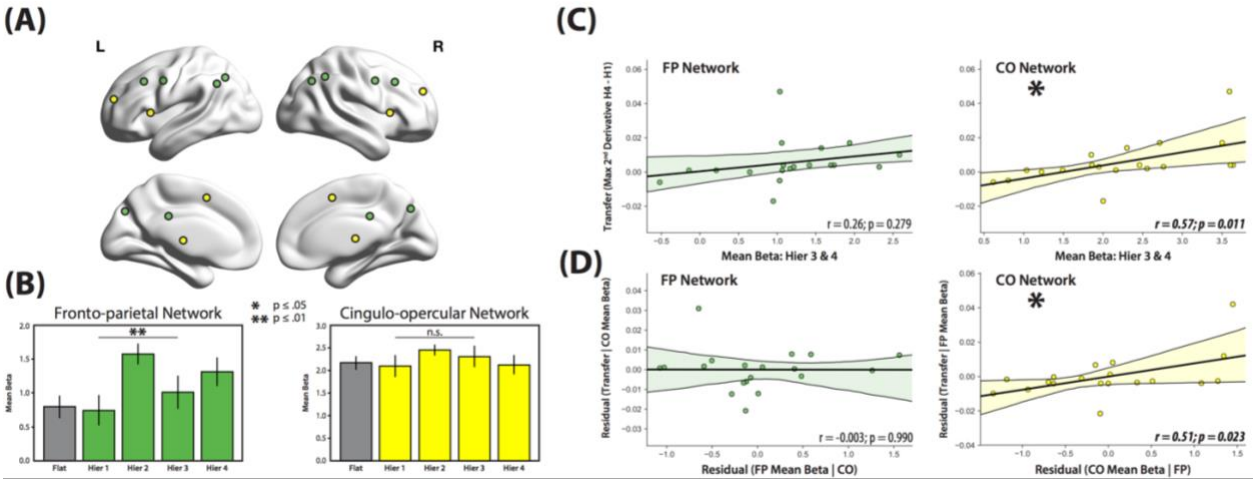


Figure 2.5 fMRI Analyses Reveal Dissociation of Behavioral Roles for FP and CO Networks. (A) Locations of regions that define the FP (green) and CO (yellow) networks defined from Dosenbach et al., 2007. FP: Bilateral frontal cortex, dorsolateral prefrontal cortex, IPL, IPS, precuneus, and midcingulate. CO: Bilateral aI/ FO, aPFC, thalamus, and dACC / msFC. (B) The contrast of the mean beta coefficients from the stimulus response phase across all respective regions in the second hierarchical block compared to the first and third blocks reveals an increase in activity during the search and discovery phase only in the FP regions. Moreover, there is a significant interaction such that the difference in activity between these blocks is greater in the FP network than in the CO network. Error bars indicate within-subject standard error of the mean. (C) The correlation of transfer with each network's mean beta coefficient from the third and fourth hierarchical blocks. (D) Regression analyses for the FP and CO networks against behavioral transfer reveal a unique role of the CO network in structure transfer. Shown are the partial-correlation coefficients from a multiple regression that accounts for the effects of both networks.

2.5 Discussion

Subjects were able to efficiently discover and exploit abstract structure during a hierarchical reinforcement learning task. During the task, subjects rapidly discovered and generalized an embedded global task structure to subsequent novel task blocks. Moreover, this generalization was supported by an increase in subjects' awareness of the specific global hierarchical structure at the start of a new block. The fMRI data revealed that multiple left lateral frontal regions were involved during task performance (prePMd, Mid-IFS, and FPC). In addition, regions within a frontal-parietal network were involved in the initial discovery of the global hierarchical structure, while regions within a cingulo-opercular network, and potentially PMd, were involved in the transfer of this structure.

Previous work on structure learning in the context of hierarchical reinforcement learning (Collins et al., 2014; Collins & Frank, 2013, 2016) has shown that subjects tend to build generalizable structures that allow for components of the stimulus (e.g. shape) to act as a higher-order context that cues rules based on other stimulus features (e.g. color). However, in contrast to previous work where stimulus-response groupings could be directly transferred, our task design prevented direct block-to-block transfer of action mappings. Instead of discovering structure that immediately informed action, such as learning one of the block-specific hierarchical policies, subjects discovered structure that informed subordinate task-set policies, as evidenced by more rapid learning in hierarchical blocks following discovery. Moreover, when a MoE model was used to derive an estimate of subjects' attention to the hierarchical shape rule at the start of the third hierarchical block, the model-derived estimate was greater than at the start of the second hierarchical block, indicating that subjects transferred and immediately applied their structural knowledge following discovery in the second hierarchical block. This demonstrates that subjects are capable of learning a higher-order representation between stimulus dimensions that can abstract away from the groupings of specific response pairings, and can then transfer this knowledge to new contexts.

Our brain imaging findings have implications regarding the functional organization of the frontal cortex in support of hierarchical learning. The lateral frontal cortex is recruited for both the learning and execution of hierarchical rules (Badre & D'Esposito, 2007; Badre et al., 2010; Badre & Nee, 2018; Collins et al., 2014; Koechlin, 2003; Nee & D'Esposito, 2016), with recruitment of more rostral regions during processing of higher levels of policy abstraction. In addition, patients with lateral frontal cortex lesions exhibit asymmetric behavioral impairments: caudal lesions impair both concrete and abstract cognitive control task performance, while rostral lesions only impair abstract task performance (Badre, Hoffman, Cooney, & D'Esposito, 2009). In tasks where hierarchical rules had to be implicitly learned, different lateral frontal regions are simultaneously involved in the search for hierarchical policy within a block (Badre et al., 2010). However, patients with pre-PMd lesions are impaired at learning the full 2nd-order policy, but not the subordinate 1st-order rules (Kayser & D'Esposito, 2013). This asymmetric functional deficit is evidence of the hierarchical organization of functions associated with these regions. Our study extends these findings by demonstrating that frontal cortex is involved in the search for a global hierarchical structure, beyond that of the block-specific 2nd-order policies, when evidence of its presence is first available. We conclude that the same hierarchical frontal cortex organization used to execute policy rules, as well as search for hierarchical relationships of varying complexity within the moment (i.e. block-specific policies), is also involved in the search for hierarchical relationships across contexts.

There existed a potential relationship between activity in the most caudal region (PMd) and the measure of transfer and implementation of global hierarchical structure, defined as the change in the maximum 2nd-derivative across blocks. The maximum 2nd-derivative captures the initial rise of the learning curve, indicating the transition from searching for higher-order rules to the resolution of 1st-order rules. Subjects are transitioning from a phase of the task where the search space of possible structures is large to one where it has become well-defined and narrow. With conflict of the 2nd-order policy resolved, all that remains is the resolution of 1st-order rules, a process linked to PMd function. It is likely that subjects who resolve the 2nd-order conflict more rapidly can then rely primarily on processes associated with PMd (i.e., linking specific

colors and textures to motor responses) for the remainder of the block, therefore facilitating performance.

Together with previous work, the current findings suggest a sophisticated coordination between motor control, rule implementation, rule discovery, and rule generalization in the service of hierarchical control, where each function incorporates knowledge of both the immediate setting (i.e. task block) and overall environment (i.e. global hierarchical structure). In simple tasks lacking contextual elements, caudal premotor regions likely resolve response competition without influence from superordinate rostral frontal regions. However, in tasks for which contextual information must be considered (e.g., abstracted hierarchical policy), rostral premotor and mid-dorsolateral regions are likely recruited to exert control over sensory-motor conflict in more caudal premotor regions (Badre et al., 2009; Kayser & D'Esposito, 2013). In settings where actions and rules are being learned, these contextual influences are likely being tested and updated via cortico-striatal interactions in response to task-based feedback signals (Badre & Frank, 2012; Frank & Badre, 2012). Thus, when a subject discovers and transfers global structure, knowledge of this structure works to restrict the search space of potential hypotheses, resulting in selective recruitment along the rostrocaudal gradient to those involved in representing the generalized known structure. Thus, multiple regions can be involved in the process of structure transfer, but specifically only those regions along the gradient necessary for the resolution of the remaining unresolved block-specific rules.

Several cortical and subcortical regions outside the lateral frontal cortex associated with behavioral transfer were identified. Subjects with greater levels of activity in regions comprising the CO network learned the block-specific hierarchical policies faster following discovery of the global hierarchical structure. Critically, this association was not found in regions comprising the FP network, suggesting that CO network activity is specifically related to the manner in which subjects maintain and implement the learned structure. Alternatively, it is possible CO activity is increased in subjects who are more engaged and attentive to the task (Sadaghiani & Kleinschmidt, 2016). We favor the former interpretation because our transfer metric indexes a difference between performance in the first, compared to final, hierarchical block, and is thus insensitive to differences between subjects who perform poorly in both phases (when it could be assumed that subjects are failing to pay attention to the current task), and those who perform exceedingly well in both phases (when it is likely that attentional engagement is greatest).

Previous work has implicated the CO network in both "task-set maintenance" (Dosenbach et al., 2006, 2007, 2008), broadly defined as the configuration of control signals required to perform any type of task, and "tonic alertness" (Sadaghiani & D'Esposito, 2015; Sadaghiani et al., 2010), or the user-driven sustained control necessary to remain prepared to process incoming information. Task-set maintenance requires that a specific structure be known to the individual – that which defines successful performance of the task – whereas tonic alertness precludes any need for a specific structural representation of the task as alertness takes the role of "nonselective disengagement" (Sadaghiani & Kleinschmidt, 2016). Thus, our findings are more consistent with a role of the CO network in task-set maintenance, although a role in tonic alertness during our task cannot be ruled out.

Whereas the CO network was uniquely related to transfer, the FP network was selectively involved in the search and discovery of the global hierarchical structure. Our findings suggest that the FP network is not only involved in the representation and integration of current task rules and response mappings, but also in the integration of previous task-relevant components. The integration of this information would likely allow for complex structured relationships to be discovered across blocks. Although the component processes of searching for and discovering abstract hierarchical structure overlap with behaviors associated with learning and navigating the explore-exploit dilemma – classically linked to regions along anterior cingulate cortex (ACC) – it is unlikely that ACC would be uniquely linked to search and discovery as additional roles associated to ACC likely occurred during the preceding and proceeding phases of the task (i.e. exploring and evaluating individual hierarchical policies in the 1st hierarchical block, representing exploitative behaviors in the 3rd hierarchical block, etc.) (Quilodran, Rothé, & Procyk, 2008; Stoll, Fontanier, & Procyk, 2016; Walton, Bannerman, Alterescu, & Rushworth, 2003). Recent studies have discovered that tasks requiring varying levels of cognitive control recruit regions along a caudal-rostral gradient in parietal cortex in a similar fashion to that found in lateral frontal cortex (Choi, Drayna, & Badre, 2018). Moreover, regions along both gradients showed mirroring patterns of functional connectivity with striatal sites, in line with previous work (Badre & Frank, 2012; Collins & Frank, 2013). Accordingly, the present results implicate a system of parallel and distributed hierarchical gradients across frontal and parietal cortex that supports the search and discovery of structure of varying complexity within and across task blocks.

CHAPTER 3

Differential contributions of static and time-varying functional connectivity to human behavior

3.1 Abstract

Measures of human brain functional connectivity acquired during the resting-state track critical aspects of behavior. Recently, fluctuations in resting-state functional connectivity patterns – typically averaged across in traditional analyses – have been considered for their potential neuroscientific relevance. There exists a lack of research on the differences between traditional "static" measures of functional connectivity and newly-considered "time-varying" measures as they relate to human behavior. Using functional magnetic resonance imagining (fMRI) data collected at rest, and a battery of behavioral measures collected outside the scanner, we determined the degree to which each modality captures aspects of personality and cognitive ability. Measures of time-varying functional connectivity were derived by fitting a Hidden Markov Model. To determine behavioral relationships, static and time-varying connectivity measures were submitted separately to canonical correlation analysis. A single relationship between static functional connectivity and behavior existed, defined by measures of personality and stable behavioral features. However, two relationships were found when using time-varying measures. The first relationship was similar to the static case. The second relationship was unique, defined by measures reflecting trialwise behavioral variability. Our findings suggest that time-varying measures of functional connectivity are capable of capturing unique aspects of behavior to which static measures are insensitive.

3.2 Introduction

Measuring activity in the human brain during a task-free "resting state" has become common as this activity is known to be spatially and temporally organized (Biswal et al., 1995). These patterns of resting-state functional connectivity (rsFC) are sensitive to numerous aspects of behavior, including cognitive performance (Chan, Park, Savalia, Petersen, & Wig, 2014; Stevens et al., 2012), age (Chan et al., 2014), and the extent of cognitive impairments (Alexander-Bloch et al., 2010; Rudie et al., 2013). Using rsFC data from the Human Connectome Project (Van Essen et al., 2013), a recent report utilized canonical correlation analysis (CCA) to reveal that rsFC and numerous behavioral measures were linked via a single mode of population covariation, providing a single inextricable link between stable functional brain organization and interindividual behavioral differences (Stephen M. Smith et al., 2015).

The majority of neuroimaging studies have investigated rsFC by assuming that it is stable across the measurement period. However, a recent emphasis has been placed on determining whether, and to what degree, rsFC systematically varies in time (Calhoun, Miller, Pearlson, & Adali, 2014). While some measurable fluctuations are likely due to noise or non-neural,

physiological signals (Duff, Makin, Cottaar, Smith, & Woolrich, 2018; Hindriks et al., 2016; Hutchison et al., 2013; Lindquist, Xu, Nebel, & Caffo, 2014; Lurie et al., 2019), there is evidence that these rapidly evolving changes have a neuronal basis (Brookes et al., 2014; Chang & Glover, 2010; de Pasquale et al., 2010; Thompson, 2018). Moreover, analysis of time-varying FC might reveal new relationships to behavior unobtainable by static analyses (Cohen, 2018; Kucyi, Tambini, Sadaghiani, Keilholz, & Cohen, 2018). There is recent evidence that fluctuations of task-based FC track aspects of cognitive control (Khambhati, Medaglia, Karuza, Thompson-Schill, & Bassett, 2018) and attention (Sadaghiani et al., 2015), suggesting that flexible network reconfiguration indexes trial-by-trial performance.

It is important to consider ways in which static and time-varying FC differ, and how these differences impact the way each modality encodes aspects of behavior. Whereas static measures provide a snapshot of the stable organization of the brain, time-varying measures index higher-order relationships between brain regions. Such measures include the degree to which functional networks vary their interconnectivity with other networks, the change in global organizational structure, and how the global FC profile transitions between different functional substates (Shine & Breakspear, 2018; Vidaurre et al., 2017). Thus, it is likely that measures of static and time-varying FC encode different behavioral features, however a precise characterization of this relationship is missing. Studies have focused on either one type of connectivity (static: Smith et al., 2015, time-varying: Casorso et al., 2019), or on specific behaviors (Rosenberg et al., 2016), but only two studies attempted to simultaneously disentangle static and time-varying FC's behavioral relevance (Jia, Hu, & Deshpande, 2014; Liegeois et al., 2019). Jia and colleagues found that time-varying measures of FC explained more variance in behaviors tracking alertness, cognition, emotion, and personality than did static FC. Liégeois and colleagues found that measures of time-varying FC tracked both task-based behavior and self-reported personality traits, whereas static measures only captured self-reported traits. Although leveraging the power of the Human Connectome Project, these studies only had access to basic measures of human behavior, lacking access to measures typically employed by cognitive neuroscientists studying working memory, cognitive control, and executive function.

In order to directly address the behavioral differences captured by static and time-varying FC, we utilized resting-state blood-oxygen-level-dependent (BOLD) data collected alongside a battery of complex behavior and personality measures. These measures ranged across working memory, executive functioning, processing speed, affect, and impulsivity. Building off Smith and colleagues (2015), we leveraged CCA to determine whether there exist modes of covariation between behavior and static, as compared to time-varying, rsFC. Static rsFC was estimated by computing a node-node correlation matrix across all regions of the brain. Time-varying rsFC was estimated by fitting a Hidden Markov Model (HMM) to the data. The HMM allowed for the characterization of, and transition likelihood between, multiple latent "states" in a data-driven fashion as fast as the modality allowed, overcoming limitations imposed by sliding-window methods (Hutchison et al., 2013). The HMM has been used to characterize brain dynamics across multiple neuroimaging modalities during rest (Baker et al., 2014; Vidaurre et al., 2017) and task (Vidaurre et al., 2018).

Using static FC, CCA revealed a single relationship primarily defined by variance in measures of personality and affect, as well as task-general behavioral features. With time-

varying FC, CCA instead revealed two (orthogonal) relationships. The first was highly similar to that found using static FC. However, the second was specific to time-varying FC, and was defined by variance in trialwise measures of reaction time to processing speed and working memory tasks, as well as measures tapping into overall processing accuracy. These results suggest that there exists meaningful information in the temporal fluctuations of rsFC patterns which can explain aspects of human behavior to which previous analytic methods have been insensitive.

3.3 Materials and Methods

Participants

Twenty-three healthy young adult participants (mean age = 28.26 years, SD = 4.52 years, 10 females) were recruited for a repeated measures study to participate in two or three sessions. Five participants were unable to attend the third study session as a result of having moved away from the state of California. As a result, only 18 participants were included in the third session (mean age = 27.67 years, SD = 4.64 years, 8 females). All participants were native English speakers, had normal or corrected-to-normal vision, and had normal hearing. Participants were excluded for any history of neurological or psychiatric disorders, use of psychotropic drugs, a history of substance abuse, or MRI contraindications. All participants provided written informed consent according to the procedures of the UC Berkeley Institutional Review Board.

Experimental Design and Procedure

Participants underwent one practice session approximately one week (mean = 6 days, SD = 2.37 days) before their first testing session. They then completed two or three identical testing sessions. Testing sessions 1 and 2 were separated by approximately one week (mean = 8 days, SD = 1.47 days), while testing sessions 2 and 3 were separated by approximately one year (mean = 399 days, SD = 28.73 days).

Each session began with two 6-minute resting-state scans in the MRI machine, in which participants were instructed to stay awake with their eyes open and fixate on a crosshair. During the first session, the resting-state scans were followed by a structural scan. Immediately after the MRI scan, participants completed two self-report questionnaires and a task outside of the scanner: a Visual Analogue Scale (VAS, McCormack, Horne, & Sheather, 1988), the Barratt Impulsiveness Scale (BIS, Patton, Stanford, & Barratt, 1995) and a Box Completion Task (Salthouse, 1996). Immediately following completion of the questionnaires and task, participants then completed four computerized cognitive tasks in counterbalanced order (different orders across participants and for each session): a Stroop task (Stroop, 1935), a digit symbol substitution task (DSST; Rypma et al., 2006), a spatial working memory (WM) task (Kuo, Yeh, Chen, & D'Esposito, 2011), and a color WM task (W. Zhang & Luck, 2008). Visual depiction of the four computerized task paradigms are shown in Figure 3.1.

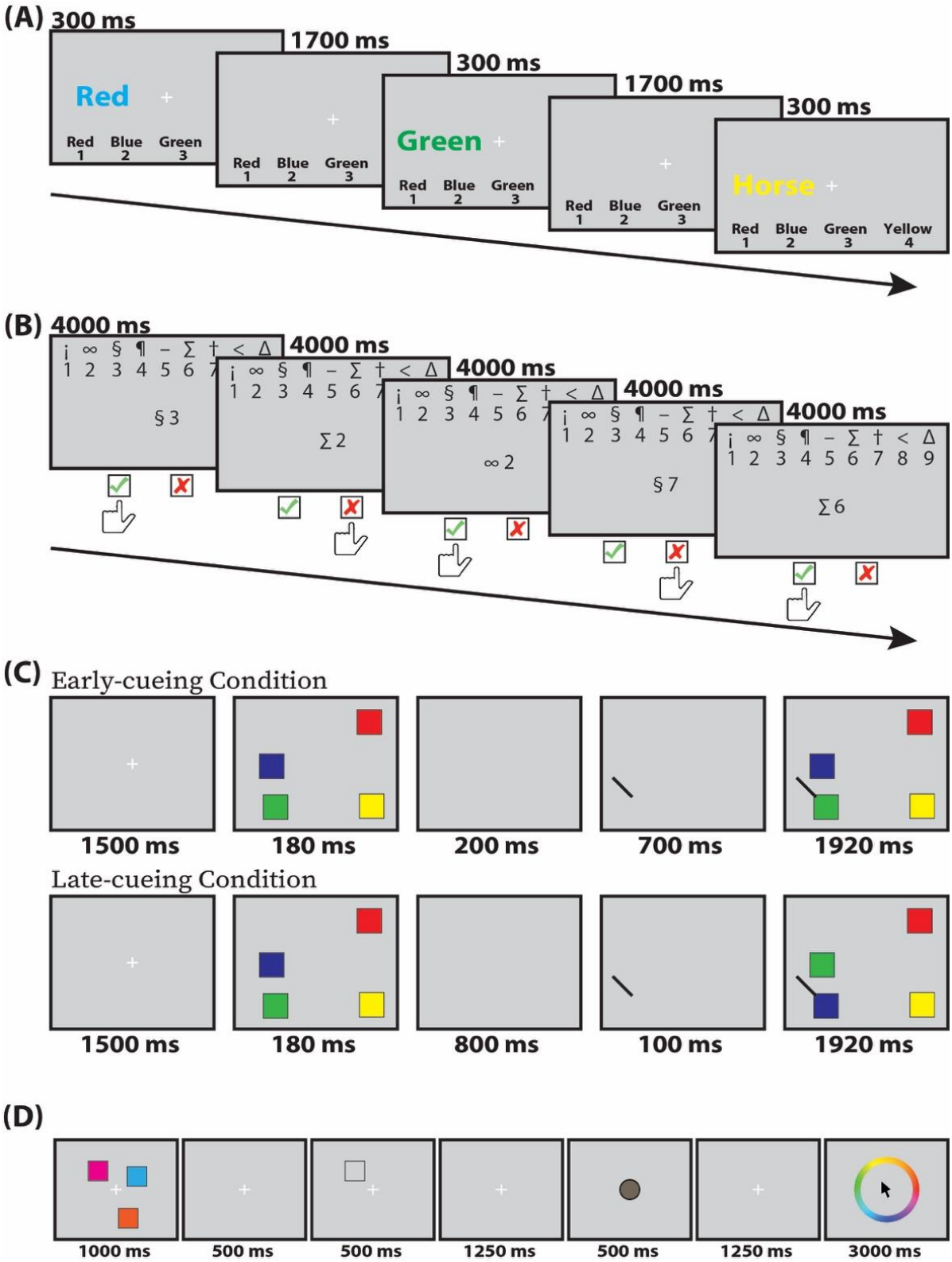


Figure 3.1 Computerized Task Paradigms. Shown are the time courses of example trials of the 4 computerized behavioral tasks. (A) Stroop task. Trial 1 is Incompatible, 2 is Compatible, and 3 is Neutral. (B) DSST. (C) Spatial WM. Both early- and late-cueing trial types are shown, with the early-cueing trial depicting a match trial and the late-cueing trial depicting a non-match trial. (D) Color WM.

The BIS is a survey that determines measures of impulsivity along a set of three sub-traits: “Attentional,” “Motor,” and “Non-planning”. The VAS has participants make a mark along a line segment in which one side represents “Not” and the other side “Extremely” for the following items: “Anxious,” “Happy,” “Sad,” “Nauseous,” “Drowsy,” “Jittery,” “Fatigued,” and “Dizzy”. Participant responses are measured as the distance (in centimeters) away from the “Not” end of the line. The Box Completion Task requires that participants use a pencil to fill in the fourth side of an open-ended square as rapidly as possible. The measure of interest is the duration of time it takes to complete 100 squares.

In the Stroop task, color words (blue, red, green, yellow) or animal names (horse, bird, cat, dog) printed in different colors (blue, red, green, yellow) were presented on the left side of the computer screen. Participants had to indicate the font color by pressing one of four buttons. For ease of task performance color-to-button mappings were presented at the bottom part of the screen throughout the duration of the experiment. Participants used the four fingers of their right hand for responding with color-to-button mappings randomly assigned to participants. Compatible, neutral, and incompatible trials were presented with equal probability. In compatible trials, color and word were the same. In neutral trials, the task-irrelevant dimension (e.g., word meaning) was not related to the task (e.g., animal names). In incompatible trials, color and word differed. Each Stroop session was ten minutes long and comprised 8 blocks of 36 trials each. The stimuli were presented for 300 ms with an interstimulus interval of 1700 ms. The measures of interest included the difference score, in milliseconds, between the median response time of correct responses to trials in which there was an incongruity between the word and color (incongruent trials: i.e., the word “RED” in blue text) and the median response time of correct responses to a trial in which the color of the text matched the word (congruent trials: i.e., the word “RED” in red text). Moreover, we also focused on the standard deviation of this response time difference, as well as the accuracy on incongruent trials. We chose not to compute a difference score for accuracy as individual differences for accuracy on congruent trials was likely to be minimal.

The DSST required that participants indicated via button press whether a presented symbol-number pair correctly matched an on-screen answer key. Nine symbols were paired with numbers 1 through 9 and the answer key was shown at the top of the screen on every trial. 140 pairs were presented in which the symbol-number pair either matched (50%), or did not match (50%), the provided answer key. Pairs were presented on screen for 4000 ms, during which the participant could indicate their response. Participants were instructed to respond as rapidly and as accurately as possible. Measures of interest included the overall accuracy, median reaction time, and standard deviation of reaction time, for match and non-match trials separately.

The spatial WM task (“Spatial WM”) required that participants initially encode and retain the color of a rapidly presented set of colored squares. The task followed a 2 (load: 2 vs. 4) x 2 (cue onset: early vs. late) design. Participants viewed an array of 2 or 4 colored squares for 180 ms prior to retaining this information over a 900 ms delay period. In the early-cue condition, a cue appeared in the location of where one of the squares had previously been after 200 ms (and stayed on screen for the remaining 700 ms). In the late-cue condition, the cue appeared after 800 ms (and stayed on screen for the remaining 100 ms). Next, participants had to indicate whether a

newly presented colored square, among an array of 2 or 4 colored squares, matched the color of the spatially cued square prior to the delay. The new array remained on screen for 1920 ms. Participants were instructed to respond as accurately and as quickly as possible. In total, participants completed 240 trials, with 60 trials coming from each condition. Measures of interest included percent accuracy, median reaction time, and the standard deviation of reaction time, across both cognitive loads, for match and non-match trials separately.

The color working memory task ("Color WM") required that participants initially encode the colors of 3 squares rapidly presented on screen for 1000 ms. Following a delay of 500 ms, a visual cue to the location of one of the squares appeared for 500 ms. After a 1250 ms delay, a distractor color appeared on screen for 500 ms. Following another delay of 1250 ms, the participants were then presented with a colorwheel for 3000 ms and were instructed to move the cursor along the wheel in a continuous fashion until the selected color matched the color of the cued square being held in memory. Participants completed 40 trials in total and were provided a 5 second break after the end of the 20th trial. Measures of interest included the median and standard deviation of reaction time and error angle (calculated as the difference in degrees along the colorwheel between the correct answer and the response provided by the participant) across all responses.

During the practice session, participants completed the four cognitive tasks so as to familiarize themselves with the tasks before the testing sessions. The purpose of this session was to minimize practice effects. The testing sessions were all identical. The final testing session was conducted on the same MRI machine as the previous sessions, but in a different location. Reliability tests ensured that MRI effects (such as signal-to-noise ratio and artifacts) were not different across the two locations.

Behavioral measures for each subject at each session were considered as separate yet dependent datapoints, and therefore no averaging across sessions occurred. Given the dependent nature of these data points, we utilize the analytic methodology from Smith et al. (2015), which accounted for familial relationships between specific subjects in the Human Connectome Project. Specifically, we performed all statistics such that permuted null distributions never shuffled labels across sessions for subjects. In other words, all sessions from any particular subject were always grouped together so as to appropriately account for within-subject variability.

Factor Analysis of the Behavioral Data

All 31 behavioral measures were included in the analyses and subjected to a factor analysis. Six measures each came from the Spatial WM task and the DSST: percent accuracy, median reaction time, and the standard deviation of reaction times for match and non-match trials. Three measures came from the Stroop task: percent accuracy on incongruent trials, median reaction time difference between congruent and incongruent trials, and the standard deviation of the reaction time difference between congruent and incongruent trials. Four measures came from the Color WM task: median and standard deviation of response error, as well as median and standard deviation of reaction times. All eight measures from the VAS were included, as well as the scores of the three sub-traits of the BIS. Lastly, the time to complete all 100 squares for the Box Completion Task was included.

We clustered the behavioral data into 8 factors using MATLAB's *factoran* function and allowed for promax oblique rotation (Figure 3.2). We labeled these factors qualitatively by observing which behavioral measures loaded highest on each factor. We chose 8 factors as it most cleanly separated tasks from one another and grouped together correlated measures.

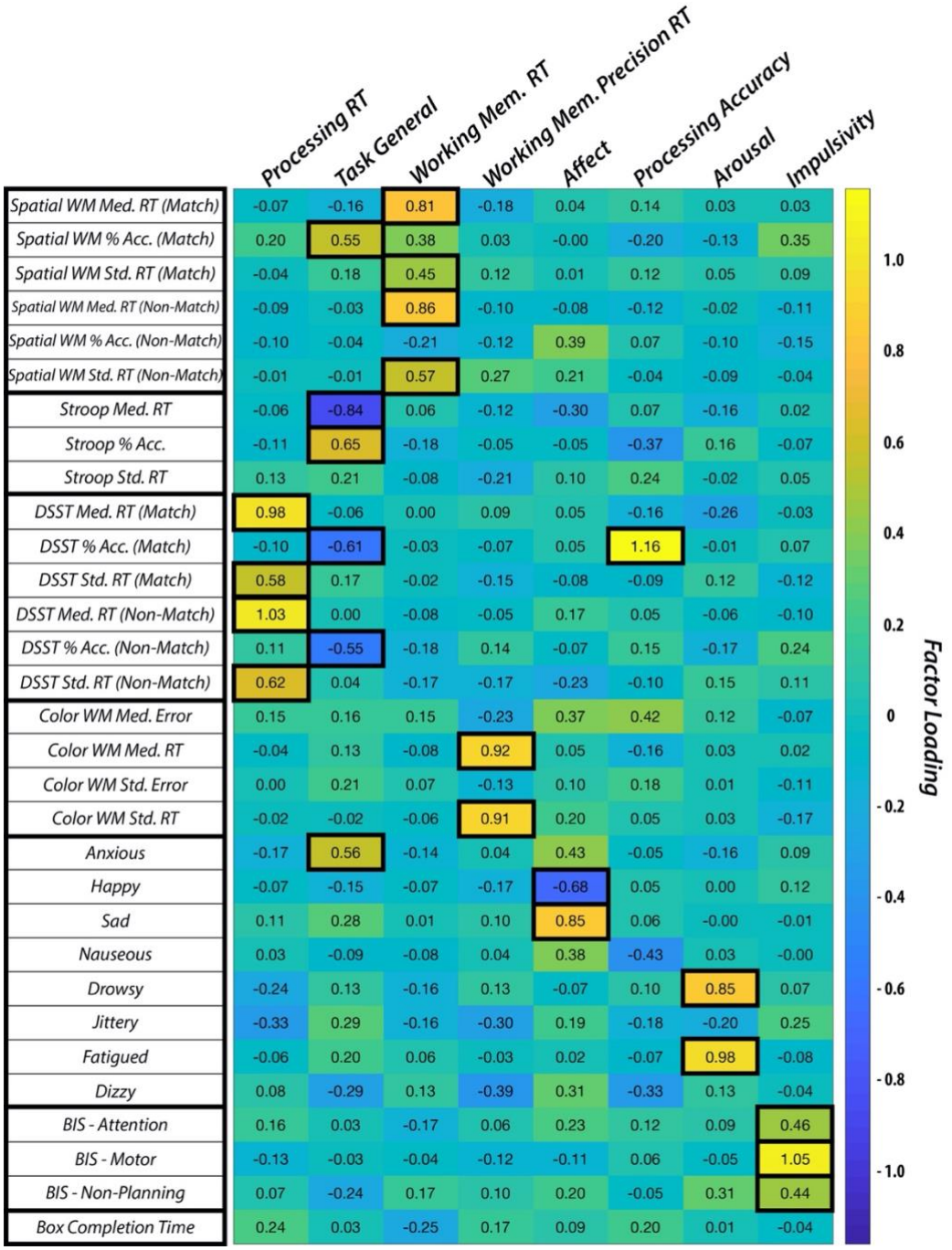


Figure 3.2 Factor Loadings in Behavioral Data. Shown are the factor loadings across the 8 derived factors for each of the 31 behavioral measures. To aid interpretability, certain values are highlighted to show the clustering of loadings within groups. Factors were assigned names based on the behavioral variables that loaded most heavily on them. RT: Reaction Time.

fMRI Data Acquisition

Imaging data were collected on a 3-Tesla Siemens MAGNETOM Trio whole-body MR scanner using a 12-channel head coil at the University of California, Berkeley Henry H. Wheeler Jr. Brain Imaging Center. Whole-brain functional data were acquired in two runs using a T2*-weighted echo-planar imaging (EPI) pulse sequence (180 volumes/run, 37 interleaved axial slices parallel to the AC-PC line, slice thickness 3.5 mm, interslice distance = 0.7 mm, TR = 2000 ms, TE = 24 ms, FA = 60°, matrix 64 x 64, field of view 224 mm). A high-resolution T1-weighted structural 3D MP-RAGE was also acquired (160 slices, slice thickness 1 mm, TR = 2300 ms, TE = 2.98 ms, FA = 9°, matrix 256 x 256, field of view 256 mm). An LCD projector back-projected a fixation cross for the resting-state scan onto a screen mounted to the RF coil.

fMRI Data Processing

Preprocessing of the imaging data were performed using fMRIPrep 1.1.4 (Esteban et al., 2018, 2019), which is based on Nipype 1.1.1 (Gorgolewski et al., 2011). The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) using N4BiasFieldCorrection (ANTs 2.2.0, Tustison et al., 2010), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped using ANTs BrainExtraction (ANTs 2.2.0), using OASIS as target template. Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, Dale, Fischl, & Sereno, 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (Klein et al., 2017). Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (MNI152NLin2009cAsym, Fonov, Evans, McKinstry, Almli, & Collins, 2009) was performed through nonlinear registration with ANTs Registration (ANTs 2.2.0, Avants et al., 2008), using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter, and gray-matter was performed on the brain-extracted T1w using fast (FSL 5.0.9, Y. Zhang et al., 2001).

For each of the BOLD runs found per participant (across all sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using mcflirt (FSL 5.0.9, Jenkinson et al., 2002). BOLD runs were slice-time corrected using 3dTshift from AFNI. The BOLD time series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time series will be referred to as preprocessed BOLD in original space, or just preprocessed BOLD. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve & Fischl, 2009a). Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. The BOLD time series were resampled to MNI152NLin2009cAsym standard space, generating a preprocessed BOLD run in MNI152NLin2009cAsym space. Several confounding time series were calculated based

on the preprocessed BOLD: framewise displacement (FD), DVARS, and three region-wise global signals. FD and DVARS were calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al., 2014). The three global signals were extracted within the CSF, the white matter, and the whole-brain masks (i.e., global signal). The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. All resamplings were performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and template spaces). Gridded (volumetric) resamplings were performed using ANTs ApplyTransforms, configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964).

Further post-processing included removal of artifactual signals from the time series data. We used recommended nuisance regression approaches based on recent processing comparisons (Ciric et al., 2017; Parkes, Fulcher, Yücel, & Fornito, 2018). We regressed out the 6 head-motion estimates, the mean white matter signal, the mean cerebral spinal fluid signal, their temporal derivatives and quadratic expansions, and the quadratic expansions of the temporal derivatives. We chose to avoid global signal regression due to (1) the known effect of introducing artefactual negative correlations into the data and (2) the increase in distance-dependent motion effects. As temporal contiguity is necessary to accurately estimate changes in FC across time, we did not apply any scrubbing techniques to our data. Last, we applied a bandpass filter from 0.01 to 0.1 Hz to the data. Mean framewise displacement of our sample was relatively low (mean FD = 0.14, range = 0.06 – 0.37) and aligned with previously analyzed samples (Power et al., 2014).

Static Functional Connectivity

In order to obtain measures of FC, we first measured the mean BOLD signal across all voxels contained within each node of our brain atlases. Cortical nodes were taken from the 400-node Local-Global atlas (Schaefer et al., 2018). 21 subcortical nodes were taken from the Harvard-Oxford atlas (Makris et al., 2006). 22 cerebellar nodes were taken from the AAL atlas (Tzourio-Mazoyer et al., 2002). Four cortical nodes in bilateral anterior temporal pole regions had to be removed from all analyses due to insufficient coverage (less than 25% of voxels contained data) in one or more participants in one or more scans. This left data from 439 nodes distributed across the entire brain.

Scans were concatenated within session, per participant, in order to increase reliability of the measured FC profile for each session. To remove spurious data differences between sessions, each session's data was standardized. FC was measured as the Pearson correlation coefficient between every node and all other nodes for which there was sufficient coverage.

Hidden Markov Model

Setup

The HMM derives brain dynamics based on BOLD time series parcellation data. The HMM assumes that the time series data are characterized by a number of states that the brain cycles through at different times throughout the scanning period (Baker et al., 2014).

At each time point *t* of brain activity, the observed time series data was modeled as a mixture of multivariate Gaussian distributions. Each one of these Gaussian distributions corresponded to a different state *k* and was described by first-order and second-order statistics (activity [μ_k] and FC [Σ_k], respectively) that can be interpreted as the activity and FC of each state. Using notation, if x_t describes the BOLD data at each time point *t*, then the probability of being in state *k* is assumed to follow a multivariate Gaussian distribution

$$P(\, x_t \mid s_t = \mathrm{k}) \sim \text{Multivariate Gaussian}(\, \mu_k, \Sigma_k \,)$$

In turn, we modeled how transitions between states took place. The basic Markovian principle that describes the transition between states assumes that the probability of the data being in state *k* at time *t* relates only to the probability of being in state *l* at time *t-1*. This can be described by the following equation

$$P(\, s_t = k \,) = \Sigma_l \, \Theta_{l,\,k} \, P(\, s_{t-1} = l \,)$$

where $\Theta_{l,\,k}$ is the transition probability from state *l* to state *k*. Taken together, the HMM infers the $P(\, s_t = k \,)$ probabilities for each state *k* and time *t* (state time courses) as well as the transition probabilities $\Theta_{l,\,k}$ and the statistics of each state (μ_k, Σ_k) that best describe the data. To make inference tractable, a variational Bayes algorithm was used that works by minimizing the Kullback-Leibler divergence between the real and the modeled data (Wainwright & Jordan, 2007).

The input time series data for the HMM was the total time series data for all participants and all sessions (for the last session there were only 18 participants). Specifically, across the three sessions and for all participants we concatenated the processed functional time series and obtained a matrix of dimensions: (360x23 + 360x23 + 360x18) × number of regions of interest (439) (Vidaurre et al., 2017). Data were standardized for each participant prior to running the HMM. To control the dimensionality in the final data matrix, a principal component analysis (PCA) dimensionality-reduction technique was applied on the concatenated time courses using 25 components (Stevner et al., 2019). Finally, the number of states for the HMM was chosen as 12. Both of these settings were similar to the previous work that introduced the use of the HMM on fMRI data (Vidaurre et al., 2017).

Inference

Running the HMM with these parameters resulted in a data matrix of dimensions (# time points x # participants) x # states. Each row represented the probability of each state being active at each timepoint for each participant. Additional quantities related to the temporal characteristics of each state could then be obtained. First, we quantified the proportion of time that an individual resided in the state during the scan acquisition (fractional occupancy; FO). Additionally, the switching rate was defined as the difference between the probability of

activating a state at time *t* and activating a state at time *t+1* summed over all states and over all time points and divided by the number of time points. The HMM also provided each state's mean activity and connectivity μ_k and Σ_k, respectively. Finally, the HMM also provided the state transition probability matrix of dimensions (# states x # states) where each matrix entry (*k*, *l*) quantified the transition probability of going from state *k* to state *l*.

An agglomerative hierarchical clustering algorithm was applied to the transition probability matrix in order to determine whether there existed a temporal structure in the data, as had previously been shown with resting-state FC data from the Human Connectome Project (Vidaurre et al., 2017). This analysis starts by classifying each data point as a separate cluster and progressively combines clusters of data at different hierarchical levels: similar data are clustered at a low level of hierarchy and less similar data are clustered at a higher level of hierarchy (Hastie, Friedman, & Tibshirani, 2009). We used the *linkage* function as implemented in MATLAB with default settings (method= 'single', distance= 'euclidean'). We regarded each identified cluster as one metastate. In turn, the metastate time courses were considered as the sum of the time courses of the individual states that comprised them. Fractional occupancy and switching rate of the metastates were calculated as in the case for single states.

In order to assess whether there existed any relationship between the derived HMM time-varying FC measures and in-scanner head motion, we first computed the mean FD across both runs for each subject in each session. There existed no difference in mean FD across sessions (all t's < 1.19, adjusted-p's = 0.679). Next, we correlated these session-specific mean FD values with each of the 17 measures derived from the HMM (i.e. fractional occupancy of each of the 12 states, fractional occupancy of each of the 3 metastates, mean switching rate across the 12 states, and mean switching rate across the 3 metastates). Given the number of statistical tests performed (17 x 3 = 51) and the related nature of the data being assessed, we applied FDR correction to our results to account for multiple comparisons. Two of the fifty-one tests survived multiple comparisons correction: Fractional Occupancy of State 1 in Session 2 ($r = 0.847$, adjusted-$p < 0.001$), and Fractional Occupancy of State 11 in Session 3 ($r = 0.734$, adjusted-$p = 0.026$). Given the lack of any consistent relationship between in-scanner head motion and our HMM measures across sessions, it is likely that our preprocessing strategy of the fMRI BOLD data appropriately corrected for motion artifact for the current study's analyses of time-varying FC.

Spatial Characterization of States

In order to spatially characterize the derived states, we thresholded the activity maps of each state to include the top 40% of both positive and negative activations. We then spatially overlapped each state with the 10 resting-state networks described in Smith et al., (2009) to obtain an overlap index for each network. The index was calculated by counting the number of voxels that were included in the thresholded map and then dividing these by the size of the resting-state network under consideration in order to account for size bias.

Canonical Correlation Analysis (CCA)

To relate the behavioral measures to static and time-varying FC we used CCA (Figure 3.3). CCA finds correlations between multidimensional data wherein potential relationships may

be present (Hotelling, 1936). This is a more principled approach compared to conducting all potential correlations and correcting for multiple comparisons. Specifically, this analysis finds maximal correlations between two sets of variables, X (n x d1) and Y (n x d2), where d1 and d2 are the number of variables used in X and Y respectively, and n is the number of observations for each variable. It produces two matrices, A and B, such that the variables U=AX and V=BY are maximally related. CCA values were obtained from the MATLAB *canoncorr* function. It is worth noting that like the PCA, this function can produce more than one mode, with each mode ranked by the covariance that can be explained between X and Y.

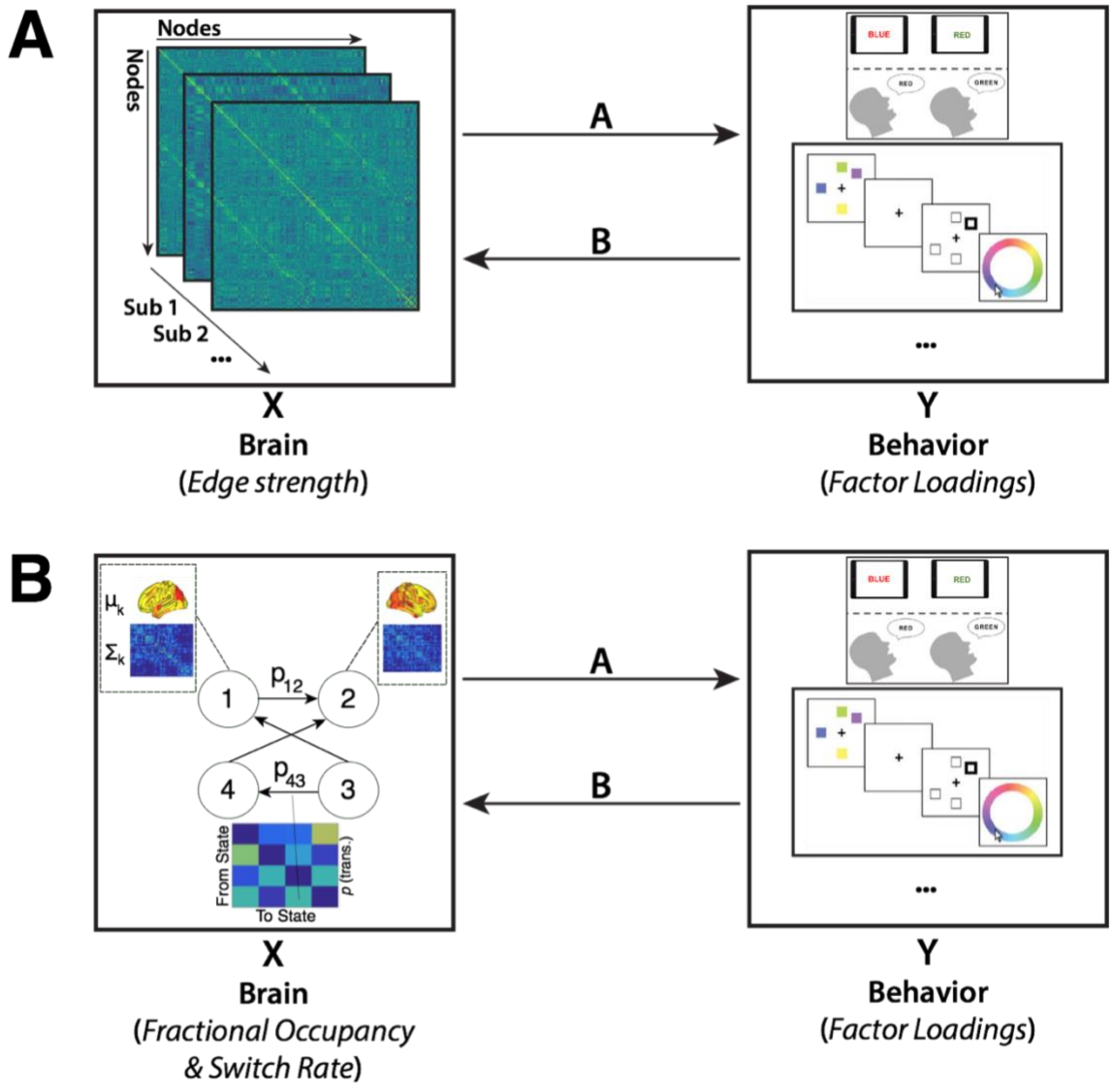


Figure 3.3 Methodology Overview**.** CCA was performed on two different datasets, which were matched for measures of behavior but differed with regard to the rsFC data included. The first CCA (A) included measures of static FC (i.e., the node-to-node connectivity strength), while the second CCA (B) included measures of time-varying FC. Measures of time-varying FC were derived by fitting a Hidden Markov Model to the BOLD time series.

We conducted two separate CCAs. First, we designated the factors of the behavioral data as Y, and the edgewise static FC strength as X (n = 96,141). In a second CCA, Y remained the same, but we varied X. Specifically, we designated X as the fractional occupancy of each HMM state (n = 12) and temporally defined metastate (n = 3), as well as the mean switching rate across states (n = 1) and metastates (n = 1) separately. As a final preprocessing step, the dimensionality of the static FC data was reduced using PCA as described in Smith et al. (2015), retaining the top 13 components. No dimensionality reduction was required for the HMM data as the number of variables was low. However, we performed an analysis of the HMM data using PCA and report that the results are highly similar to the case when PCA is not employed (see *Validation of CCA Analysis* below).

Statistical significance of the CCA analyses was estimated as follows. We calculated 10,000 permutations of the rows of X relative to Y, respecting the within-participant structure of the data, and recalculated the CCA mode for each permutation in order to build a distribution of canonical variate pair correlation values (i.e. <U,V>). By comparing the outcome from the CCA of the true data to the shuffled data, we found that each mode of covariation discovered with the true data was highly significant ($p < 1/10,000$). In addition, a cross-validation approach was adopted in order to assess the robustness of the discovered mode(s) (as described in Smith et al., 2015). Across 1,000 runs, we ran CCA on a randomly selected set of 80% of the data, respecting the within-participant nature of the data, and stored the resultant U and V. We then estimated the mode on the held-out 20% of data and determined the significance of the estimated mode employing the same permutation significance testing procedure as before. These estimated modes were found to be highly significant, with the correlation between the derived canonical weight vectors in the test dataset being very robust (replicating the results from Smith et al., 2015).

Post-hoc correlations of the values of X (Y-respective) with the columns of the significant mode U (V-respective) were used to quantify the contributions (positively or negatively) of each behavioral measure with the CCA mode. In other words, we quantified the extent to which the Y variables were loaded/weighted on the CCA mode. There is no clear cutoff at which one finds a significant correlation value and thus correlation values are reported in isolation.

Validation of CCA Analysis

We validated the identified CCA modes by comparing outcomes across a range of behavioral factors (behavior) and FC principal components (static FC, time-varying FC). The number of behavioral factors ranged from 1 to 9, while the number of static FC principal components ranged from 1 to 20 and the number of time-varying FC principal components ranged from 1 to 17. For the static case, we ran CCA on each combination and stored the resulting post-hoc correlations for each behavioral measure (i.e., with respect to FC), and computed the Pearson correlation between these values across all 180 combinations (Figures 3.4, 3.5). For combinations that included two or more behavioral factors, we found that the discovered canonical covariate modes were highly similar, with Pearson correlations tending to be very highly positive (i.e., greater than r = 0.90) as well as very highly negative (i.e., less than r = -0.90). This bimodal distribution at the extremes of the correlation range indicates that the

discovered modes were highly preserved in structure (i.e., the same behavioral measures loaded highly). We determined the optimal combination (i.e., 8 behavioral factors, 13 FC PCs) by selecting either (A) the most significant canonical covariate pair (i.e. U x V), or (B) in cases where multiple pairs had the same maximal 1 / 10,000 permutation significance value, determining if combinations were highly similar after a certain number of factors or components were included, and then taking the smallest number of factors and components that produced this outcome, restricted by those that had a significant permutation value.

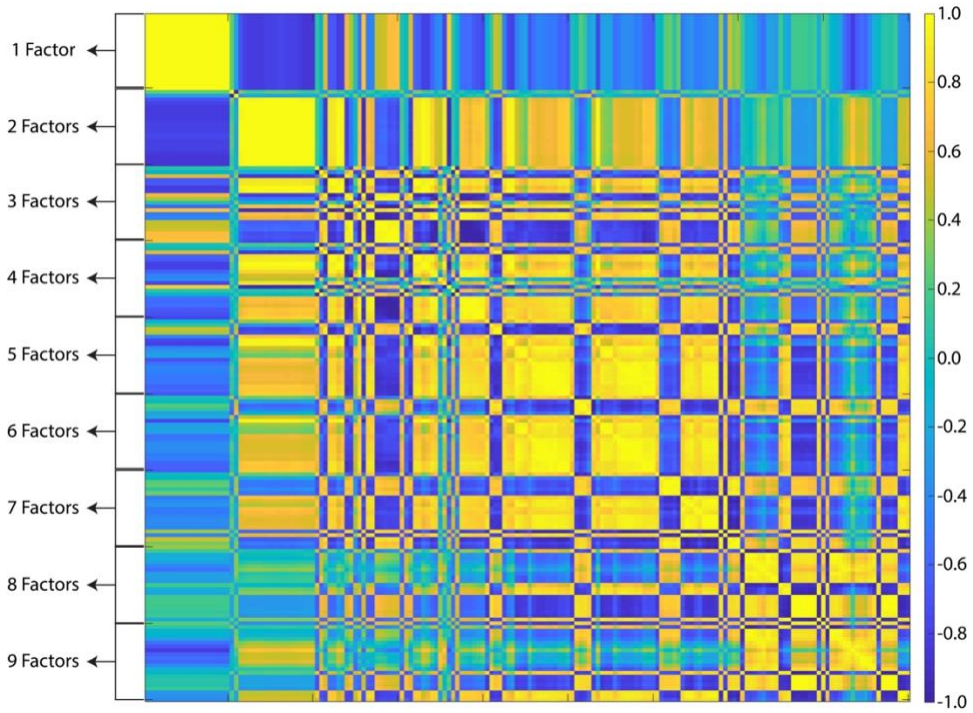


Figure 3.4 Validation of Static Functional Connectivity CCA. Similarity between modes of covariation are plotted across varying numbers of behavioral factors (1-9) and edge strength principle components (1-20). Similarity was measured as the Pearson correlation coefficient between each mode's vector of 31 behavioral post-hoc correlations. The plot is organized such that for each CCA run, connectivity data in varying principal component space is grouped together within chunks of behavioral data in a fixed factor-space. For example, the first row represents CCA being run on behavioral data in 1-factor space and connectivity data in 1-PC space; the 10th row represents CCA being run on behavioral data in 1-factor space and connectivity data in 10-PC space; the 25th row represents CCA being run on behavioral data in 2-factor space and connectivity data in 5-PC space, etc. Essentially, only one mode is found across all PCs when in 1- and 2-factor space, with the discovered modes being identical within factor space, and almost exact inverses of each other across factor spaces. Modes discovered from CCA run on data in 3-factor space and above, across all PCs, almost entirely represent the same relationship as each other.

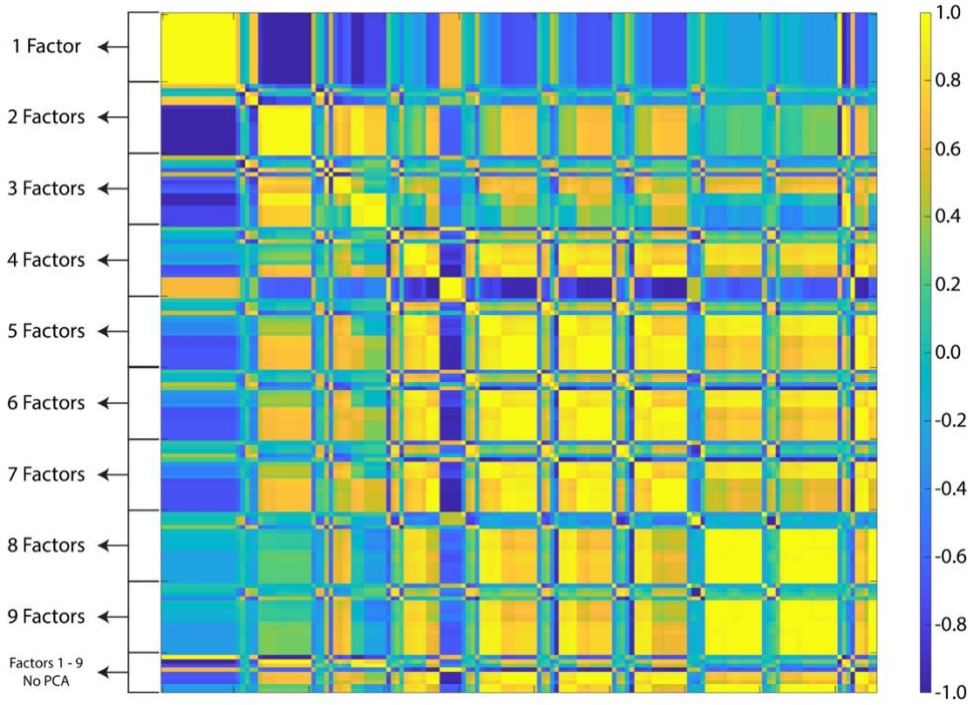


Figure 3.5 Similarity between modes of covariation are plotted across varying numbers of behavioral factors (1-9) and edge strength principle components (1-17, capped by number of HMM variables included). Similarity was measured in the same way as for the static FC CCA validation procedure using Pearson correlation. Unlike in the static FC case, we additionally assessed whether the discovered modes differed as a function of whether PCA was, or was not, run on the HMM data. The final 9 rows represent the modes discovered from CCA run on behavioral data in 1- through 9-factor space when PCA was not performed on the HMM data. As can be seen, the modes discovered when CCA was run on the raw HHM data almost perfectly match those found when PCA was run on the HMM data. A corresponding plot for post-hoc correlations computed from modes greater than 1 is inappropriate as we had no *a priori* knowledge that more than 1 significant mode of covariation would be discovered, and thus only the first mode of covariation can be compared across datasets.

3.4 Results

Factor Analysis

Brain and behavioral data were obtained as described in the Methods. We used factor analysis to reduce the 31 behavioral measures to 8 factors (Figure 3.2). The first factor, referred to as "Processing Reaction Time", had DSST median and standard deviation reaction time measures for both match and non-match trials loading highly positively. The second factor was referred to as "Task General" because it contained a mixture of measures across multiple tasks, with positive loadings from the Spatial WM task percent accuracy (match trials), the Stroop task percent accuracy, and "Anxious" on the VAS, and negative loadings on the Stroop task median reaction time and the DSST percent accuracy (both match and non-match trials). The third factor, referred to as "Working Memory Reaction Time," had the Spatial WM task median and standard deviation reaction time measures, for both match and non-match trials, loading highly positively. The fourth factor, referred to as "Working Memory Precision Reaction Time," had two Color WM task measures loading highly positively: median and standard deviation of reaction time.

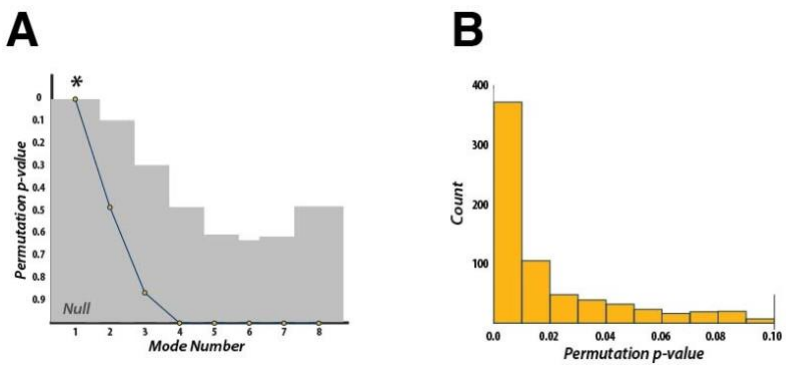
The fifth factor, referred to as “Affect,” had the VAS measures “Sad” and “Happy” loading highly positively and negatively, respectively. The sixth factor, referred to as “Processing Accuracy,” had only the DSST percent accuracy on match trials loading highly positively. The seventh factor, referred to as “Arousal,” had high positive loadings for both the “Drowsy” and “Jittery” VAS measures. Finally, the eighth factor, referred to as “Impulsivity,” included high positive loadings of all three BIS measures.

The first (“Processing Reaction Time”), third (“Working Memory Reaction Time”), and fourth (“Working Memory Precision Reaction Time”) factors all contain measures of both the median and standard deviation of reaction time across the DSST, Spatial WM, and Color WM tasks, respectively, and therefore reflect aspects of within-task stability (median reaction time) and within-task variability (standard deviation of reaction time). In contrast, the second (“Task General”) and sixth (“Processing Accuracy”) factors only contain task measures of accuracy and/or median reaction time, and thus only reflect aspects of within-task stability. Lastly, the fifth (“Affect”), seventh (“Arousal”), and eighth (“Impulsivity”) factors all contain measures that reflect the personality and mood of the participant.

Canonical Correlation Analysis: Static Functional Connectivity

CCA was used to find a mode of population covariation between behavior and static FC. The CCA included the behavioral data in 8-factor space, as well the static rsFC data in 13-principal component space, based on the validation we performed (see the "Validation of CCA Analysis" section of the Methods for details). The CCA revealed a single mode of covariation between these two datasets (Figure 3.6). To assess the statistical significance of the discovered modes of covariation, we followed the permutation and cross-validation procedure as outlined in Smith and colleagues (2015; see “Canonical Correlation Analysis (CCA)” section in Methods, and Figure 3.6A, B).

We used post-hoc correlations between the discovered mode and the behavioral factors to evaluate the contribution of each factor to the mode, with respect to the static FC data. This mode was defined by highly positive weights for the “Affect” ($r = 0.69$), “Task General” ($r = 0.54$), and “Working Memory Precision Reaction Time” ($r = 0.30$) factors, and a highly negative weight for the “Impulsivity” ($r = -0.30$) factor. All other factors had correlation values below an absolute value of 0.11. These results indicate that static connectivity might encode more general behavioral and personality features rather than information that may relate more to task, or trial-specific, behavior.



Mode x Factor Relationships

Factor	Correlation
Affect	0.69
Task General	0.54
Working Mem. Precision RT	0.30
Working Memory RT	0.11
Arousal	0.05
Processing Accuracy	0.04
Processing RT	0.02
Impulsivity	- 0.30

Figure 3.6 Canonical Correlation Analysis – Static Functional Connectivity. (A) CCA can discover as many modes of covariation as the lowest rank of each dataset (i.e., 8 behavioral factors). Statistical significance was found only for the first discovered mode. (B) Additional cross-validation of the discovered mode revealed that the first mode was statistically robust across the majority of the 1,000 folds. (C) Post-hoc correlations for the discovered mode and the 8 behavioral factors revealed that measures of "Affect" and "Impulsivity", as well as a "Task General" factor, dictated the structure of the mode. RT: Reaction Time.

Canonical Correlation Analysis: Time-Varying Functional Connectivity

We next assessed whether any relationships existed between time-varying FC and behavior. To quantify the time-varying FC profile in each participant we fit the resting-state BOLD data with a HMM. This model works by finding relevant states and their associated spatial (activity, connectivity) and temporal (fractional occupancy, switching rate) characteristics (see the "Hidden Markov Model" section in the Methods). After fitting the HMM, we identified 12 states that were representative of brain dynamics across all participants (Figure 3.7). Previous work has shown that the transition probabilities between HMM states derived from resting-state data is structured (Vidaurre et al., 2017). Specifically, there are certain sets of states, or "metastates", that are more temporally coherent than others. In other words, if a participant visits a state within one metastate they are more likely to stay within that metastate compared to transitioning to another metastate. Hierarchically clustering the transition probability matrix resulted in three main clusters. One included two states, another included nine states, and the third included a single state. These results are similar to those found previously with the Human Connectome Project dataset (Vidaurre et al., 2017), indicating that even with our comparatively small sample size, we could reliably estimate brain dynamics. For completeness, we included all twelve states in our analysis; however, our results remained unchanged when we excluded the state that failed to cluster with the other states.

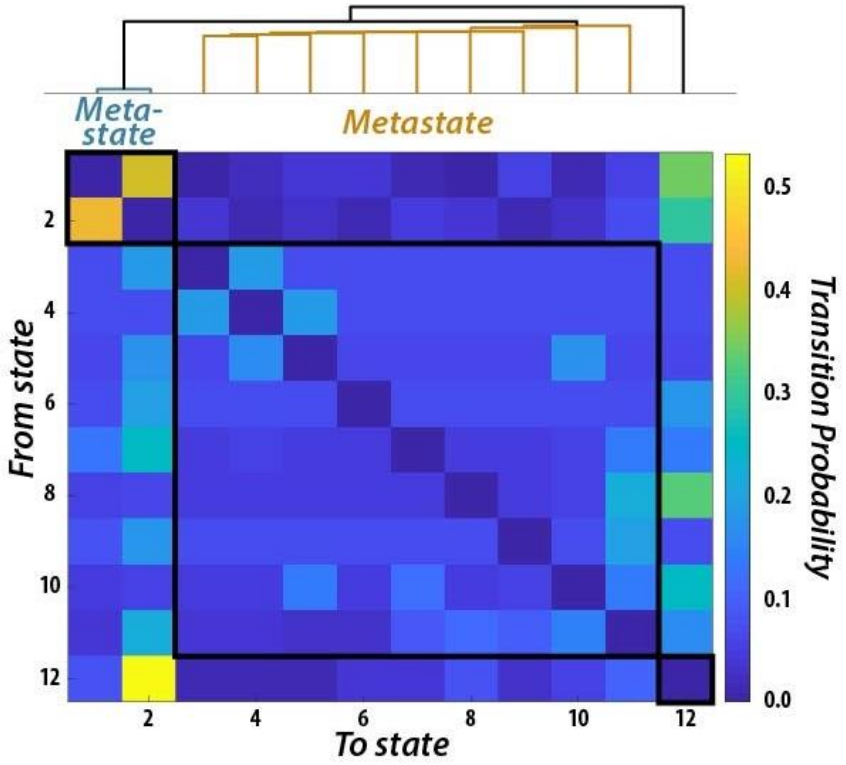


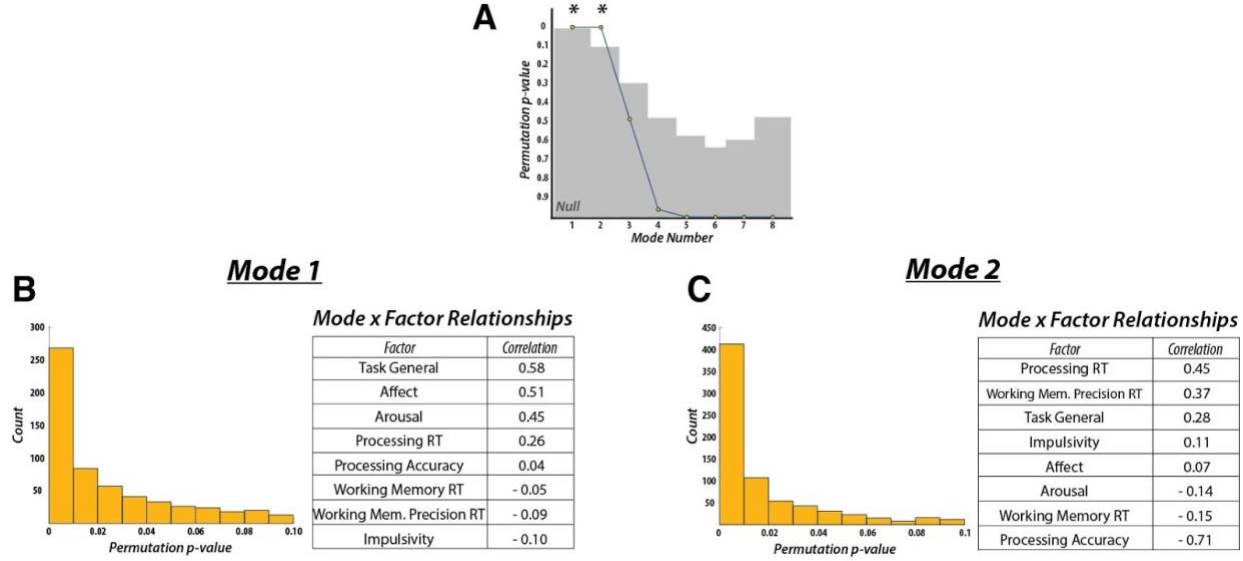
Figure 3.7 Metastates Resulting from the Temporal Clustering of Brain Dynamics. Probability, across all participants, of transitioning from one state to another. Clustering of the 12-state transition probability matrix revealed a temporal hierarchy wherein groups of states preferentially transitioned within groupings compared to across groupings. Two groupings contained multiple states (i.e. "metastates"), while one state was clustered only with itself.

Next, we used the fractional occupancy (i.e., time spent in each state) of each state and metastate, as well as the mean switching rate between states and metastates (n = 17 in total), as input into a CCA to determine the relationship between time-varying FC characteristics and the behavioral factors (n = 8, see Methods for description of selection and validation process). We found two significant CCA modes using the same permutation testing and cross-validation procedure as employed for static FC (Figure 3.8A).

The first mode was defined by positive weights for "Task General" (r = 0.58), "Affect" (r = 0.51), "Arousal" (r = 0.45), and "Processing Reaction Time" (r = 0.26) factors, showing a similar pattern to the mode obtained from static FC (Figure 3.8B). Specifically, "Task General" and "Affect" loaded highest, while "Impulsivity" (r = -0.10) loaded most negatively (although its loading was greatly reduced compared to the previously discovered static mode). All other loadings fell below an absolute value of 0.09.

The second mode exhibited different behavioral weights when compared to the first time-varying mode. Here, "Task General" (r = 0.28), "Affect" (r = 0.07), and "Arousal" (r = -0.14) factors had substantially lower weights. Instead, "Processing Reaction Time" (r = 0.45) and "Working Memory Precision Reaction Time" (r = 0.37) factors loaded most highly on the positive end, while the "Processing Accuracy" (r = -0.71) factor loaded most negatively (Figure 3.8C). All remaining factors had weights below an absolute value of 0.15.

Similar to a previous analysis on the differentiable contributions of static and time-varying FC (Liegeois et al., 2019), we found that time-varying FC, while showing some similar relationships to behavior as static FC, could also distinguish relationships with more task-based measures of behavior. However, by using more specific measures of working memory (i.e. match-to-sample vs. free recall, accuracy vs. reaction time), task processing, and cognitive control, we were additionally able to determine that the second time-varying CCA mode distinguished unique behaviors associated with task performance. Specifically, the mode was defined by a separation (i.e., a positive-negative split in post-hoc correlations) between reaction time and accuracy, thus revealing within-task effects that previously had not been interrogated.



Factor	Correlation
Task General	0.58
Affect	0.51
Arousal	0.45
Processing RT	0.26
Processing Accuracy	0.04
Working Memory RT	- 0.05
Working Mem. Precision RT	- 0.09
Impulsivity	- 0.10

Factor	Correlation
Processing RT	0.45
Working Mem. Precision RT	0.37
Task General	0.28
Impulsivity	0.11
Affect	0.07
Arousal	- 0.14
Working Memory RT	- 0.15
Processing Accuracy	- 0.71

Figure 3.8 Canonical Correlation Analysis - Time-Varying Functional Connectivity. (A) CCA performed on measures of time-varying FC revealed two significant modes of covariation. Results of the cross-validation procedure and post-hoc correlations between (B) mode 1 and (C) mode 2 revealed that both modes were highly robust (assessed across 1,000 folds) and were sensitive to different sets of behavioral features. Whereas mode 1 largely matched the mode discovered with static measures of FC, mode 2 was instead sensitive to task- and trial-specific measures of behavior. RT: Reaction Time.

To further characterize each state obtained from the HMM, we overlapped their spatial profiles with those of canonical rsFC networks (Smith et al., 2009). Qualitatively, we found that the two-state metastate overlapped with two distinct task-positive networks (i.e., fronto-parietal and somatomotor networks; Figure 3.9). The nine-state metastate overlapped with a larger variety of networks, including the default mode, executive, and visual networks (Figure 3.9). Unthresholded spatial maps of each of the 12 states are shown in Figure 3.10.

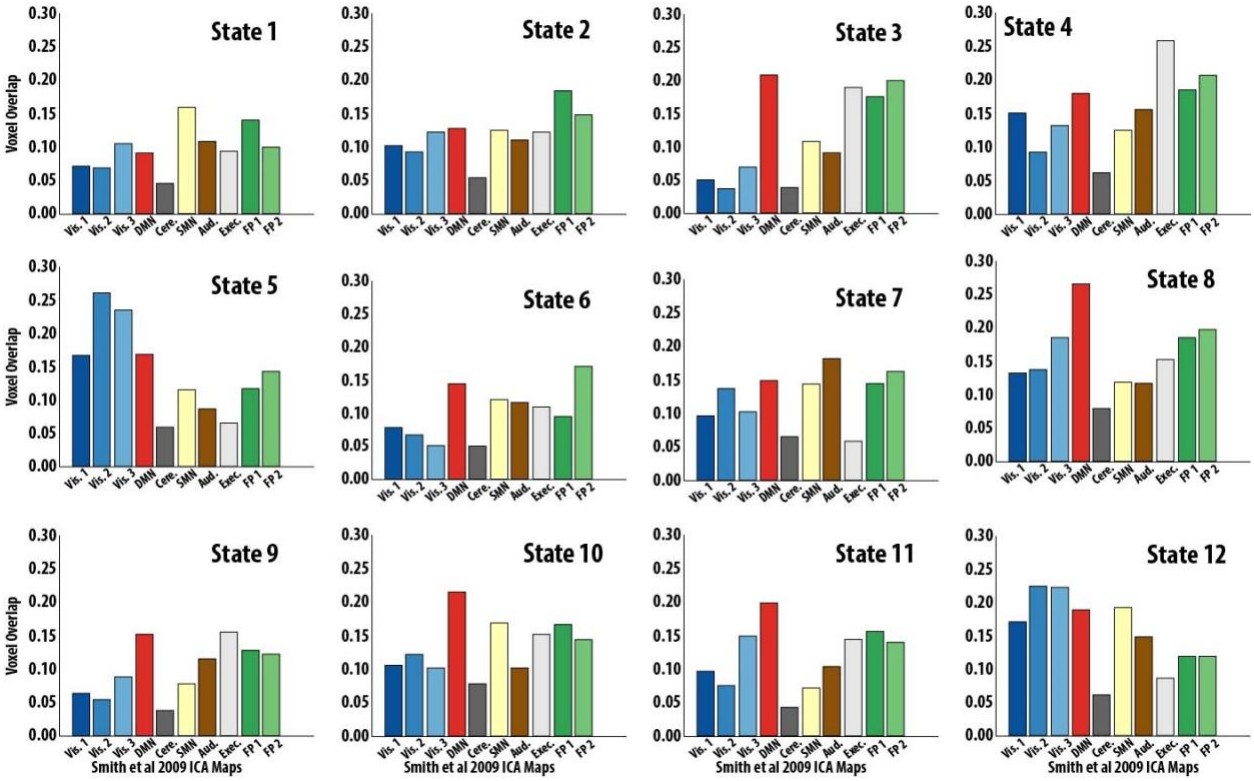


Figure 3.9 HMM State Activation Map Overlap with Resting-State Networks. Voxel overlap proportion for each HMM-derived state to the ten resting-state ICA maps from Smith and colleagues (2009). Ordering of states matches that of Figure 3.7. Specifically, states 1 and 2 clustered together in one metastate, states 3-11 in another metastate, and state 12 clustered alone.

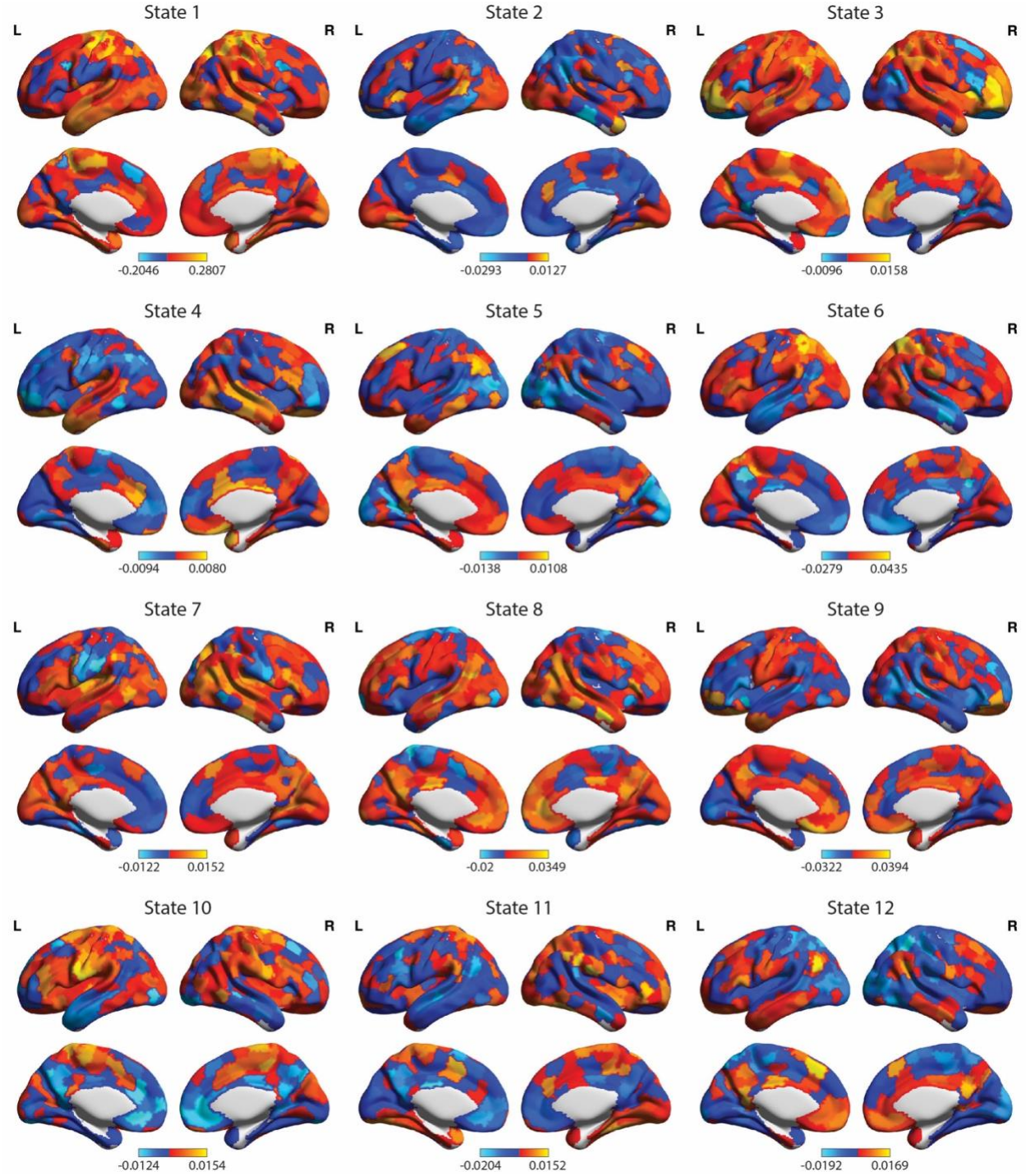


Figure 3.10 Spatial profile of HMM States. Shown are the mean activation profiles (positive and negative) for each of the 12 HMM states. States 1 and 2 formed one metastate, states 3-11 formed another metastate, and state 12 clustered by itself.

3.5 Discussion

Using CCA, we investigated the relationship between complex measures of human behavior and both static and time-varying rsFC. We found a single CCA mode between the behavioral measures and static FC. In contrast, we found two CCA modes relating behavior and time-varying FC. Of these two modes, the first one resembled its static counterpart, while the other appeared to be distinct in that it was more sensitive to measures of task-specific behavioral variability. We thus argue that time-varying FC describes nuanced brain-behavior relationships distinctly from that which is captured by static FC.

Measures of static FC typically consider average FC over a prolonged period of time (e.g., several minutes of an fMRI scan) and have been used extensively to study the functional organization of the brain during rest and task performance (Cohen & D'Esposito, 2016; Cole, Bassett, Power, Braver, & Petersen, 2014). We used nodal bidirectional FC edge strength quantified across the entire scan in a CCA to relate the brain's intrinsic static functional organization to behavior. The CCA revealed a significant relationship between these measures and our behavioral factors. Measures of affect and impulsivity determined the main positive and negative directions of this mode, respectively. To a lesser extent, the positive direction of this mode was also characterized by a "Task General" factor. This factor represents accuracy measures derived from tasks sensitive to working memory (Spatial WM), cognitive control (Stroop), and processing speed (DSST). This factor also contains a high loading for the median reaction time measure from the Stroop task. The "Task General" factor is thus most similar to previous analyses that have used data from the Human Connectome Project where behavioral measures include median reaction times either in isolation or multiplexed with task-specific performance values. These results indicate that static FC likely tracks participant-level personality measures present during the scanning session (e.g., affective state). Unsurprisingly, static connectivity is also sensitive to measures of task performance that likely characterize stable behavioral features of the individual (i.e., general, multi-task performance, including working memory). As such, these results largely confirm the findings of previous studies on static rsFC's predictive power in regard to certain measures of human behavior.

However, it has been shown that FC, including measures from resting-state protocols, is likely a dynamic process whereby fluctuations in regional connectivity occur rapidly (Lurie et al., 2019). Given the rate at which they occur, these fluctuations might better encode behavioral information reflecting ongoing cognitive demands, as compared to a general characteristic which would remain stable over the span of minutes, hours, or days. Previously, Casorso and colleagues (2019) assessed a similar, albeit broader, hypothesis by extracting time-varying rsFC components and submitting these to a CCA along with behavioral measures collected in the Human Connectome Project. Two modes of covariation were found between their time-varying components and behavior; however, no analysis of static FC was made against which to compare. One mode was largely defined by positive post-hoc correlation values for vocabulary comprehension and working memory, and negative values for prosocial behaviors. The second mode was defined by positive post-hoc correlation values for visuospatial orienting and emotional processing, and negative values for inter- and intrapersonal processing and wellbeing. Although a critical step forward in the analysis and validation of time-varying FC's relevance to human behavior, this study did not address the nature of how time-varying FC relates to behavior

in a unique manner compared to static FC measures. Specifically, the Human Connectome Project behavioral measures used preclude the ability to measure processes that likely vary from trial to trial, as task-specific measures of reaction time only reflect the median, and not trialwise variability. In our experiment, we recorded behavioral measures that separately tracked processes related to stable (e.g., accuracy) vs. time-varying (e.g., reaction time) aspects of behavior to better assess our hypothesis. Although reaction time variability and accuracy measures never loaded highly together on any factor, it should be noted that our two measures of reaction time (i.e., median and standard deviation) loaded together on the three factors representing working memory and processing speed reaction times. Even so, the high loading of reaction time variability measures in these factors represents a novel behavioral measure compared to previous reports using publicly available datasets.

Using measures of time-varying FC calculated from fitting a HMM to our rsFC data, we investigated whether CCA would reveal modes of population covariation sensitive to measures of behavioral variability. Our analysis resulted in two significant modes. One mode largely resembled the mode discovered with static measures of FC. The primary difference between these modes is that this time-varying FC mode carried a highly positive weight for measures of drowsiness and fatigue, potentially reflecting a sensitivity of time-varying FC to neural and physiological correlates of arousal (Patanaik et al., 2018).

Whereas one of the time-varying modes reflected a largely similar, but not identical, behavioral profile as the static FC mode, the other time-varying mode reflected a more unique behavioral profile. High positive weights were associated with response time measures for tasks that assessed working memory and processing speed, while a strong negative weight was found for the measure of accuracy on the processing speed task. Characterized in part by measures of trial-by-trial response variability, this mode's positive end potentially reflects a greater sensitivity to behavioral dynamics that occur on a more rapid timescale compared to what static FC is likely sensitive. In addition, the separation of measures of response variability and overall response accuracy, especially within the same task, reveals that time-varying FC is likely capable of disentangling unique behavioral components within the same task. Although our static FC mode did show some sensitivity to a measure that captures response variability, the distinction between stable and time-varying components of behavior was not present as is seen in our second time-varying mode. Overall, it is possible that this time-varying mode captures the relationship between brain dynamics and the measures of trial-by-trial behavioral variability within complex measures of human behavior.

The manner by which time-varying fluctuations in rsFC relate to independent measures of human behavior remains unresolved. It is known that the spatial organization of functional connections changes in response to different tasks compared to rest (Cohen & D'Esposito, 2016; Cole et al., 2014). Specifically, inter-network connectivity is more predominant during tasks that require flexible cognition (i.e., working memory) compared to more rudimentary tasks such as finger-tapping. Moreover, a previous report found that measures of global network integration and within-network connectivity (i.e., participation coefficient and module degree, respectively), when assessed in a time-varying manner, varied throughout the performance of tasks and tracked the cognitive complexity of the task demands (Shine et al., 2016). Thus, one hypothesis as to how resting dynamics relate to behavior is that the dynamic interactions within and between

these networks observed during tasks can be recapitulated during periods of wakeful rest. However, it should be noted that the dynamic interactions that occur during task performance are likely more constrained than during rest due to the confined cognitive context required by task performance. Resting-state dynamics can serve as a "baseline" repertoire that can potentially index the extent to which FC reconfigures during task and, in turn, track behavioral performance (Liegeois et al., 2019). It will be crucial for future studies on the behavioral relevance of time-varying FC to assess this possibility.

It is also important to emphasize the spatio-temporal signature of these time-varying network interactions and what it means for behavioral performance. Methods such as the HMM investigate brain dynamics with high temporal resolution, thus extending previous methods showing reconfiguration of connectivity between different task blocks (Cohen, 2018). For example, Vidaurre and colleagues used a HMM to show how a motor task drives reconfiguration of large-scale networks on a timepoint-by-timepoint basis showing that task execution happens at faster timescales that had been previously undetected when interrogated using sliding window methods (Vidaurre et al., 2018). Regarding the spatial profile of the current HMM states, a visual and quantitative assessment of their overlap with canonical rsFC networks (Smith et al., 2009) suggested that our metastates had distinct spatial profiles. We identified a nine-state metastate spanning multiple networks including fronto-parietal, executive, default-mode, and visual networks. Integration of the "task-positive" and "task-negative" networks has been observed during motor tapping and autobiographical planning, suggesting a more mutually compatible role than previously believed (Fox et al., 2005), one that can facilitate goal-directed cognition (Braga, Sharp, Leeson, Wise, & Leech, 2013; Spreng, Stevens, Chamberlain, Gilmore, & Schacter, 2010; Vatansever, Menon, Manktelow, Sahakian, & Stamatakis, 2015). On the other hand, the two-state metastate we identified, characterized by a more constrained spatial profile of fronto-parietal and somatomotor networks, potentially reflects networks specific to task execution. The differences in spatial topography of the two-state versus nine-state metastates may provide insight regarding the different behavioral relationships we found with static versus time-varying FC. The flexible interaction of activity across each metastate's respective individual states might allow for the encoding of information to which static measures are insensitive. Although static measures are capable of reflecting multi-network interactions, they are incapable of tapping into the specific temporal patterns through which these network interactions occur. Further investigation of the spatial patterns of these states is needed.

In conclusion, the current study demonstrates that static and time-varying FC are differentially associated with behavior. We argue that via integration across multiple networks at different temporal scales, time-varying FC is associated with both trial-by-trial and stable behavioral measures, while static FC is associated with participant-level personality measures and measures of stable task-general performance. These results demonstrate that it is important for future studies to look at both the static and temporal aspects of FC to more fully delineate the behavioral contributions of each.

CHAPTER 4

The role of physiological and neural measures in tracking baseball-related cognitive abilities

4.1 Abstract

The practice of scouting for and training baseball talent is undergoing a revolution as newly developed analytics and measurement tools are beginning to eclipse traditional practices. Quantifiable measures derived from sport-agnostic modalities, such as those derived from human biology, hold the potential to index aspects of human behavior and future potential that can be leveraged by sports psychologists and data scientists to better identify and hone sport-specific talent. The current report collected baseline measurements of neurophysiological functioning in order to assess their unique and shared ability to predict cognitive abilities crucial for successful baseball performance. Cardiac and respiratory timeseries data was collected from human subjects (male and female) undergoing 20 minutes of resting-state functional magnetic resonance imaging. Subsequently, subjects performed four computerized tasks measuring multiple forms of reaction time and inhibitory control. All baseball-related cognitive skills were tracked by at least one physiological or neural measure. Skills related to basic reaction time and accuracy, as well the intersection of inhibition accuracy and reaction time, were tracked by neural measures. Physiological measures were able to track basic reaction time and intersection of inhibition accuracy and reaction time, but not basic accuracy. In addition, physiological measures were able to separately track inhibition accuracy and reaction time. These results provide evidence that the collection of multiple forms of human biological functioning can aid in the scouting and training of baseball talent.

4.2 Introduction

Assessing current talent and scouting for future potential is common across sports, especially baseball. The practice has traditionally relied on deriving game-specific metrics and observing hard-to-define intangible qualities. However, recent approaches are leveraging more general and objective measures to assess both ability and potential. Some of these measures are evolutions of traditional count-based in-game statistics (e.g., On-base Plus Slugging percentage), while some represent player-specific ability (e.g., pitched-ball rotation, batted-ball exit velocity, etc.). There is also a focus on player-specific measures obtained outside the context of the game itself. These include biophysiological and psychological measures such as heart rate (HR, Duran, Tapiero, & Michael, 2018; Kennedy & Scholey, 2000) and heart rate variability (HRV, Hansen, Johnsen, & Thayer, 2003; Thayer, Hansen, Saus-Rose, & Johnsen, 2009), cognitive skill (e.g., reaction time (RT, Burris et al., 2018), and most recently, patterns of neural activity measured with electroencephalography or functional magnetic resonance imaging (fMRI, Muraskin et al., 2016; Sherwin, Muraskin, & Sajda, 2012; Sie et al., 2019).

It is highly unlikely that any one measure can predict overall performance in baseball. Therefore, it is advantageous to obtain a comprehensive repertoire of measures from each player in order to better understand which measures track which skills. Previous reports have found that measures of human physiology, such as signatures of cardiac and respiratory functioning, track aspects of human behavior. Variability in baseline measures of HR has been shown to track the processing of informative emotional cues (Duran et al., 2018), to predict error-based RT in a working memory updating task (Hansen et al., 2003), and lower HR was associated with enhanced performance on a demanding mental arithmetic task (Kennedy & Scholey, 2000). In addition, measures of HRV, known to be a robust indicator of autonomic nervous system functioning (Thayer et al., 2009), have been shown to track numerous aspects of cognitive control and attention (Hovland et al., 2012; Park & Thayer, 2014; Park, Vasey, Van Bavel, & Thayer, 2013). There have also been reports of these measures being used to assess baseball-specific behaviors, specifically for HRV (Cornell et al., 2017) and HRV biofeedback training (Strack, 2003). Last, the rate of respiration has been found to increase under cognitive load (Grassmann, Vlemincx, von Leupoldt, Mittelstädt, & Van den Bergh, 2016) and stabilize during periods of sustained attention (Vlemincx, Taelman, De Peuter, Van Diest, & Van Den Bergh, 2011). Therefore, measures of physiological functioning are likely to index numerous aspects of behavior and cognition.

The resting human brain shows patterns of activity that are spatially and temporally organized when measured using fMRI (Biswal et al., 1995). The organization of these activity patterns represents a highly modular community structure, such that activity patterns of brain regions within communities are highly correlated compared to the patterns observed from regions belonging to different communities (Sporns & Betzel, 2016). While the overall functional organization of the human brain is largely preserved across individuals (Zuo et al., 2017) – with exceptions seen in cases of neurological pathology (Alexander-Bloch et al., 2010) – individual differences due to age (Chan et al., 2014), emotion (Cisler et al., 2013), and cognitive ability (Stevens et al., 2012) have been observed across multiple scales of this structure. Moreover, there exists substantial evidence that this organization is plastic (Chan et al., 2014; Sadaghiani et al., 2015) and can index training-related improvement potential following specific forms of behavioral training (Gallen & Esposito, 2019). As such, one's baseline functional brain organization provides an ability to measure the broadscale characterization of one's current, and potentially future, behavioral profile. Currently, there are only scant reports on how brain activity is related to baseball ability and performance (Muraskin et al., 2016; Sherwin et al., 2012; Sie et al., 2019), with no reports focusing on functional brain organization.

Measures such as HRV and cognitive skill can be obtained without hassle on the playing field or in the locker room, while other measures such as those derived from fMRI require specialized settings and protocols. It is important to assess the shared and unique predictive power of these measurements in order to understand the limits and advantages of each measurement modality. Therefore, the current study collected measures of functional brain organization alongside measures of resting cardiac and respiratory functioning. These measures were used to predict cognitive skills assessed across a range of computerized tasks designed to tap skills necessary for successful baseball performance. These skills included both simple and multi-choice RT, as well as multiple measures of inhibitory control. Additionally, subjects provided feedback on the manner in which they performed these tasks, allowing us to assess

potential differences related to task strategy. Data were collected from the general population as we sought to determine how measures of physiology and brain structure tracked baseball-specific skills agnostic of expertise or a prolonged history of training in the sport.

Overall, we found that individual differences in all baseball-related cognitive abilities were tracked by the recorded measurements, with some being associated only with measures of physiology or functional brain organization, while others were tracked by both physiological functioning and functional brain organization. In addition, one set of skills showed separate relationships to biology based on the subject-specified strategy used during the task. These results suggest that the collection of multi-modal datasets proves advantageous in assessing skills related to successful baseball performance.

4.3 Materials and Methods

Human Subject Details

Ninety-two healthy right-handed subjects (range: 18 – 30 years; mean = 20.02; SD = 2.63; 37 males) with normal or corrected-to-normal vision participated in the study at the University of California – Berkeley. All behavioral analyses presented here include data from the 74 subjects for whom we obtained a complete dataset of acceptable quality (i.e., 1 anatomical scan, 2 functional scans, at least 10 minutes of cardiac recordings, at least 10 minutes of respiratory recordings, completion of all 4 behavioral tasks, and completion of post-study questionnaire; range: 18 – 30 years; mean = 19.76; SD = 2.57; 30 males; see below for definitions of modality-specific quality control thresholds). Two subjects were removed due to excessive motion during both resting state scans. Two subjects failed to complete both resting-state scans. Fourteen subjects did not have physiological data recorded during either resting-state scan due to either hardware failure of the recording devices or corruption of data when transferring data from the recording computer to storage. All research protocols were approved by the Committee for Protection of Human Subjects at the University of California, Berkeley. Informed and written consent was obtained from all subjects prior to participation. Behavioral data was collected in collaboration with NeuroScouting (Somerville, MA). I, along with all committee members, declare no financial conflicts of interest. NeuroScouting does not retain the rights to the data presented herein.

Experimental Design and Statistical Analyses

Instructions and Protocol

First, we obtained experimental consent, and confirmed both study and MRI scanner eligibility, prior to the start of data collection. Subjects were then escorted to the MRI scanner suite where the scanning protocol was explained to them. Prior to placing the subject inside the scanner, the cardiac and respiration recording devices were attached to the subject (see "*Physiological Data Acquisition*" for details). Next, subjects were placed inside the scanner and the lights were turned off in the scanner suite. We then collected the anatomical scan, followed

by two back-to-back 10-minute functional resting-state scans. Upon completion of the final resting-state scan, subjects were removed from the scanner and escorted to a separate room where they were placed in a testing booth with a laptop. The subjects then completed four computerized tasks (see below for details, Figure 1). Last, the subjects completed a questionnaire that probed their experience with the 4 computerized tasks they had just performed. Subjects received compensation at a rate of $20 per hour.

Task 1: Simple RT

In the *Simple RT* task, subjects had to hit the spacebar as fast as possible following the appearance of a baseball image on screen. Feedback was provided in the form of points that started at 0 and accrued throughout the task. The number of earned points appeared on screen immediately after a response was made. More points were awarded for faster, correct responses. A leaderboard showing their within-task point total compared to all other subjects was displayed in the bottom left of the screen and updated after every trial. Subjects completed 50 trials in total.

Task 2: rRT

In the *rRT* task, subjects had to hit the spacebar as fast as possible following the appearance of a baseball image with vertically-oriented laces ("Go" trials). Subjects were instructed to inhibit their response if the baseball image was rotated 45˚ to either the right or the left ("No-Go" trials). Feedback was provided in the form of points that started at 0 and accrued throughout the task. Points were calculated based on both speed and accuracy, with faster, correct responses earning the most points. The number of earned points appeared on screen immediately after a response was made. A leaderboard showing their within-task point total compared to all other subjects was displayed in the bottom left of the screen and updated after every trial. Subjects completed 80 trials in total: 38 Go trials, 42 No-Go trials.

Task 3: Rapid Adjust

In the *Rapid Adjust* task, subjects took on the role of a batter and viewed a ball travel "towards" them on screen as though it was being thrown at them by a pitcher standing atop a pitcher's mound. Trials began with the appearance of a white ball above the pitcher's mound. Immediately upon its appearance, the ball would then rapidly move "towards" the batter, visualized as a ball moving downwards along a curved trajectory, thus visually replicating an incoming pitch. On half the trials, the ball remained white. During these trials, the subject had to time their response (hitting the spacebar) such that the ball was located within the confines of a 3-dimensional box at the bottom of the screen (i.e., the "strike zone") at the moment the spacebar was hit. On the other half of trials, the ball would unexpectedly turn red after a varied amount of time prior to entering the strike zone. Subjects were instructed to inhibit their response if the ball turned red at any point. Feedback was provided in the form of points that started at 0 and accrued throughout the task. Points were awarded at a fixed rate for correct rejections of red-ball trials, while they were based on timing and accuracy (i.e. responding when the ball is in the center, compared to the periphery, of the strike zone earns more points) for white-ball trials. The number of earned points appeared on screen immediately after a response was made. A leaderboard showing their within-task point total compared to all other subjects was displayed in the bottom

left of the screen and updated after every trial. Subjects completed 156 trials in total: 78 white-ball trials, 78 red-ball trials.

Task 4: Paced RT

In the *Paced RT* task, subjects had to press the correct button, of four possible options, as fast and as accurately as possible after the appearance of a baseball image on screen. Each trial began with four empty circles presented side-by-side in a line across the center of the screen. After a variable amount of time, an image of a rapidly-rotating-in-place baseball appeared in the center of one of the four empty circles. From left to right, the circles were mapped to the D, F, J, and K buttons, and subjects had to place their left middle and index fingers on D and F, respectively, and their right index and middle fingers on J and K, respectively. For example, if the baseball appeared in the circle that was 2nd from the left, they had to press the F button with the left index finger. Feedback was provided in the form of points that started at 0 and accrued throughout the task. Points were awarded based on speed and accuracy, with faster, correct responses earning the most points. Unbeknownst to the subjects, a single 12-button sequence of responses repeated itself twenty times. Thus, subjects completed 240 trials in total.

Post-study Questionnaire:

The questionnaire was completed with pencil and paper and contained six questions. The first question was, "*On a scale of 1-5, how fatigued/tired were you as you ran through today's modules? (1 - Very Tired; 3 - Normal; 5 - Very Alert)*." The second question was, "*How do you feel like you did in rRT? Did you feel like you were weighting speed over accuracy or were you able to balance those throughout the session?*" The subject responses were scored by the experimenter as fitting into one of the following three groups: "Accuracy," "Speed," "Balanced." The third question was, "*What was the hardest part of Rapid Adjust? Did you feel like you were having more trouble hitting on white pitches or holding off on red pitches?*" The subject responses were scored by the experimenter as fitting into one of the following three groups: "White," "Red," or "Both." The fourth question was, "*Did you feel like you got better over the course of the Paced RT session?*" The subject responses were scored by the experimenter as fitting into one of the following two groups: "Yes," "No." The fifth question was, "*There are two versions of Paced RT that are administered equally across participants. For half the participants, the sequence is determined in a RANDOM manner. For the other half, there was a sequence that is administered on REPEAT. Do you think you were in the Random or Repeat group?*" The subject responses were scored by the experimenter as fitting into one of the following two groups: "Random," or "Repeat." The sixth question was, "*On a scale of 1-5, how confident are you that you were in that group (1 - Not confident at all; 3 - Somewhat confident; 5 - Very confident)?*"

Statistical Analyses of Behavioral Data

Analysis of behavioral data included the use of factor analysis to reduce the dimensionality of the behavioral data. In addition, analyses of covariance (ANCOVA) were used to investigate the effect of questionnaire responses on behavioral factor scores. Independent-

samples t-tests were used to perform follow-up means comparisons between questionnaire-defined subgroups.

Statistical Analyses of Physiological and fMRI Data

Behavioral associations between measures of physiology and measures derived from fMRI recordings were performed using multiple linear regression, as well as parametric bivariate regression for questionnaire-defined subgroup-specific behavioral associations.

Factor Analysis of the Behavioral Data

All 18 behavioral measures were included in the analyses and subjected to a factor analysis. Two measures came from the *Simple RT* task: (1) mean RT and (2) the standard deviation of RT. Five measures came from the *rRT* task: (1) the mean RT to correct "Go" trials, (2) the standard deviation of RT to correct "Go" trials, (3) the coefficient of variation for RT to correct "Go" trials (i.e., 1 – [Std. Dev / Mean]), (4) dprime, and (5) the stop rate (i.e., percentage of "No-Go" trials where the subject correctly inhibited their response). Three measures came from the *Rapid Adjust* task: (1) dprime, (2) the stop rate (i.e., percentage of red-pitch trials where the subject correctly inhibited their response), and (3) the point of subjective equality (i.e., "PSE": the duration of time the ball can stay white before turning red where the subject has a 50% chance of correctly inhibiting their response). Eight measures came from the *Paced RT* task: (1-4) the mean accuracy across trials 1-60, 61-120, 121-180, and 181-240, and (5-8) the mean RT across trials 1-60, 61-120, 121-180, and 181-240.

We clustered the behavioral data into 5 factors using MATLAB's *factoran* function and allowed for promax oblique rotation (Figure 2). We chose 5 factors as the Eigenvalues decreased dramatically after the 5th factor across multiple model attempts, as well as Eigenvalues being less than or approximately equal to 1 (representing a factor that only explains as much variance as a single behavioral measure) after the 5th factor. We labeled these factors qualitatively by observing which behavioral measures loaded highest on each factor.

fMRI Data Acquisition

Whole-brain imaging was performed at the Henry H. Wheeler Jr. Brain Imaging Center at UC Berkeley using a Seimens 3T Trio MRI scanner using a 32-channel head coil. Functional imaging data was acquired across two identical 10-minute runs with a gradient-echo echo-planar pulse sequence using a multi-band acceleration factor of 2 (TR = 1,500ms, TE = 24.8ms, flip angle = 55˚, array = 70 x 70, 50 slices, voxel size = 3.0mm isotropic). T1-weighted MEMPRAGE anatomical images were collected as well (TR = 2,530ms, TE = 1.64ms, flip angle = 7˚, array = 256 x 256, 176 slices, voxel size = 1mm isotropic). Subject's head movement was restricted using foam padding. An LCD projector back-projected a blank gray-colored screen for the resting-state scans onto a screen mounted to the RF coil.

fMRI Data Preprocessing

Results included in this manuscript come from preprocessing performed using *fMRIPrep* 1.4.1 (Esteban et al., 2018), which is based on *Nipype* 1.2.0 (Gorgolewski et al., 2011).

Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.2.0 (Avants et al., 2008), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 5.0.9, Zhang, Brady, & Smith, 2001). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* [Fonov, Evans, McKinstry, Almli, & Collins, 2009, TemplateFlow ID: MNI152NLin2009cAsym].

Functional data preprocessing

For each of the 2 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. The BOLD reference was then co-registered to the T1w reference using flirt (FSL 5.0.9, Jenkinson & Smith, 2001) with the boundary-based registration (Greve & Fischl, 2009b) cost-function. Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 5.0.9, Jenkinson, Bannister, Brady, & Smith, 2002). BOLD runs were slice-time corrected using 3dTshift from AFNI 20160207 (Cox & Hyde, 1997). The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD time-series were resampled into standard space, generating a *preprocessed BOLD run in ['MNI152NLin2009cAsym'] space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al., 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi, Restom, Liau, & Liu, 2007). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 5% variable voxels within a mask

covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the *k* components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

Further post-processing included removal of artifactual signals from the time series data. Following recent nuisance regression comparisons (Ciric et al., 2017; Parkes et al., 2018), we regressed out the 6 head-motion estimates, their temporal derivatives and quadratic expansions, the quadratic expansions of the temporal derivatives, and the first 6 aCompCor components. We chose to avoid global signal regression due to the effect of introducing artefactual negative correlations into the data. Last, we applied a bandpass filter from 0.009Hz to 0.1Hz to the data. Subjects were discarded from analysis if more than half of their volumes, across both resting-state scans, were associated with a framewise displacement (FD, Power, Barnes, Snyder, Schlaggar, & Petersen, 2012) value greater than 0.20mm.

Functional Connectivity and Graph Measures

In order to obtain measures of functional connectivity, we first measured the mean BOLD signal across all voxels contained within each node of the 400-node Local-Global cortical brain atlas (Schaeffer et al., 2018). We then computed the pairwise functional connectivity between all nodes using the Pearson correlation coefficient. Correlation matrices were computed on each 10-minute resting-state scan individually and then averaged together to produce a single mean correlation matrix per subject. Negative edges were then removed by being set to a value of 0. Modularity was then measured with the Brain Connectivity Toolbox (Rubinov & Sporns, 2010) using the *community_louvain* function, using the Yeo 7-network (Yeo et al., 2011) allegiance values as the initial community affiliation vector, and setting *gamma* to 1. In addition, we computed the mean functional connectivity within each of the 7 Yeo networks by taking the average value of all positive edges within each network.

Physiological Data Acquisition

We collected physiological data during scanning by using the BIOPAC physiological monitoring system (http://www.biopac.com). Cardiac functioning was measured using a photoplethysmograph that was secured to the subject's right index finger, or in the case signal quality was too low, their right middle finger. Subjects were instructed to leave their right hand in a single position following fitting so as to avoid contamination of the signal with movement-related artefacts.

We measured the subjects' respiration with a pneumatic pressure sensor placed below the sternum, along the right side of their ribcage. The sensor was held in place with tape as well as being compressed against the body with an elastic band. Signals were checked for consistency following initial setup, and again prior to the start of scanning. Both respiration and cardiac signals were recorded alongside analog TTL signals generated by the MRI scanner. Physiological data was recorded consistently throughout the scanning procedure at a sampling rate of 125Hz.

Physiological Data Preprocessing

Physiological data was isolated for the two 10-minute resting-state scanning periods and processing individually. Data was indexed as occurring between the first sample following the first TTL pulse and the last sample recorded prior to the final TTL pulse. All data collected prior to the first TTL pulse and after the last TTL pulse was not used for analysis. Data was then standardized between values of 0 and 1. Next, a Butterworth bandpass filter was applied to each 10-minute window using the SciPy (https://www.scipy.org/) functions *signal.butter* and *signal.lfilter*. For the respiratory data, the bandpass filter ranged from 0.25Hz to 0.50Hz. These numbers were derived by inspecting the power spectra of the recorded data and identifying the range which included the maximum power across subjects. For the cardiac data, the bandpass filter ranged from 1Hz to 6Hz. These numbers were also derived based on inspection of the power spectra across subjects, as well as including a high enough low-pass filter so as to preserve the secondary peak seen in a typical cardiac pulse.

A peak detection algorithm (Eli Billauer, http://billauer.co.il/peakdet.html) was then used to identify the timepoints of local maxima in the data. For respiration, this was associated with the point of deepest inhalation for the current breath. For the cardiac signal, this peak represented the systolic peak amplitude of the beat. Visual inspection of all detected peaks was performed by the experimenter to identify potential outliers or improperly labeled peaks in the data. The indices of these peaks were recorded so as to compute the instantaneous per-minute rate and standard deviation of the instantaneous per-minute rate for both cardiac and respiratory signatures, as well as the root mean square of successive differences (RMSSD) measure of HRV (Shaffer & Ginsberg, 2017). Instantaneous rate measures were computed by convolving the array of peaks (i.e., a 75,000-element vector [600s x 125Hz = 75,000 samples] with a value of 1 at the timepoint of the detected peaks and 0's elsewhere) with a uniform window 1 minute (i.e. 600 samples) in length. Convolution occurred with maximal overlap and therefore the resultant timeseries was 9 minutes in length, comprising 67,500 samples. This new timeseries was then down sampled to 1Hz. To compute the mean HR and mean BR for each subject, the two down-sampled timeseries were temporally concatenated and then averaged over to obtain a single value. The same process was followed to obtain the standard deviation of the instantaneous HR

and BR of each subject. To obtain the RMSSD measure of HRV, we calculated the successive time difference between heartbeats intervals (i.e., the difference in milliseconds between successive beat-beat intervals), squared those difference values, computed the average, and then computed the square root of that average (Shaffer & Ginsberg, 2017).

One subject had low cardiac signal (i.e. systolic peaks could not be detected separately from diastolic peaks or overall noise) during their first resting-state run, and therefore their HR measures were computed using only the 10 minutes of data acquired during their second resting-state scan. Another subject had low cardiac and respiratory signal during their second resting-state scan (i.e. systolic peaks could not be detected separately from diastolic peaks or overall noise, and BR peaks could not be detected at all), and therefore both HR and BR measures were computed using only the 10 minutes of signal recorded during their first resting-state scan. For both subjects, metrics computed from single-run data fell within normal ranges for mean and standard deviation compared to subjects with 20 minutes of data. All other subjects for whom physiological data was collected had sufficiently high signal quality during both of their resting-state scans.

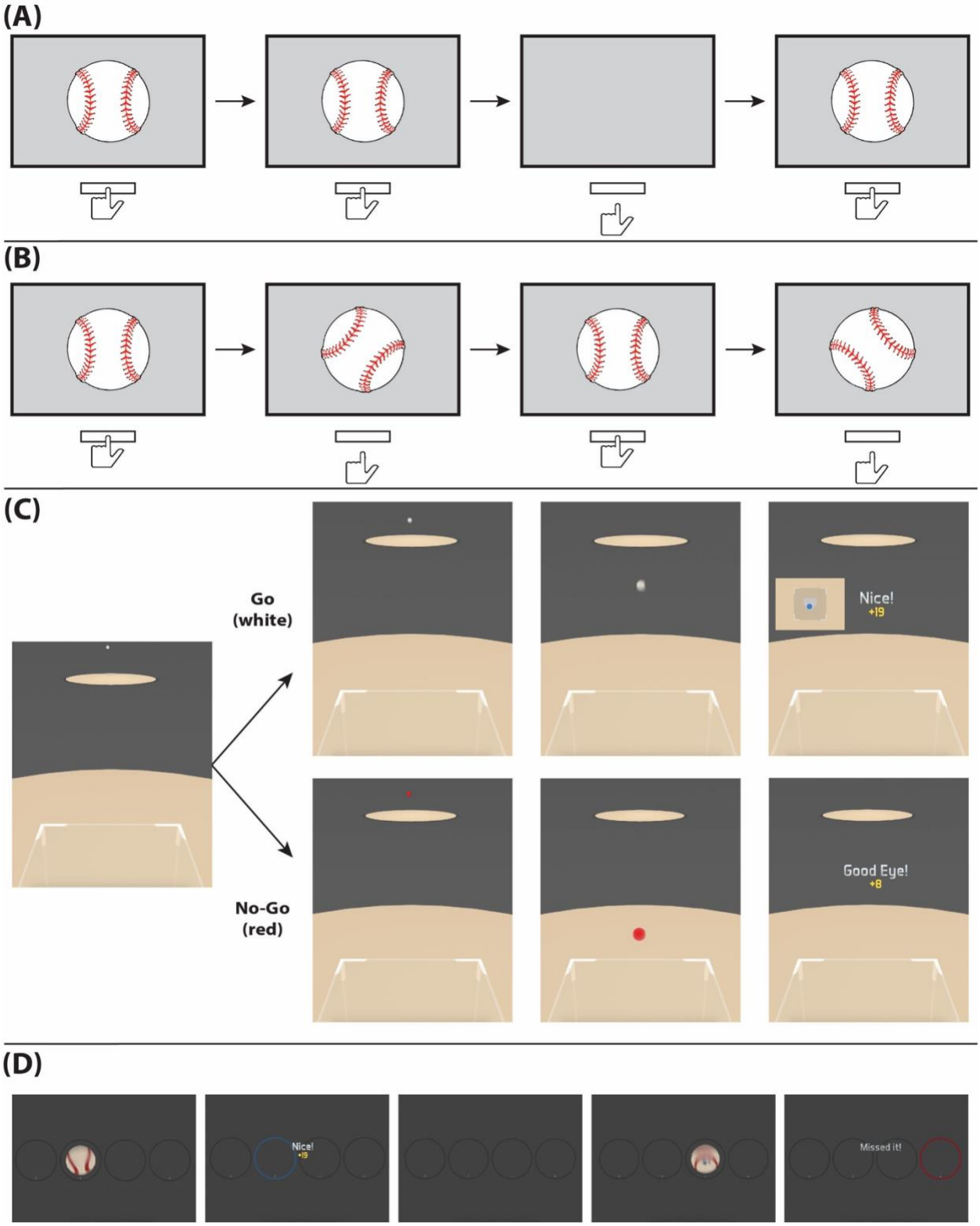


Figure 4.1 Computerized behavioral tasks. (A) *Simple RT*. (B) *rRT*. All trial types depicted. (C) *Rapid Adjust*. Both Go (white ball) and No-Go (red ball) trials are depicted with successful outcomes. (D) *Paced RT*. Both correct and incorrect response outcomes are shown.

4.4 Results

Factor Analysis

We used factor analysis to reduce the 18 behavioral measures to 5 factors (Figure 2). The first factor was referred to as "*Paced RT Reaction Time*" because it had very high positive loadings for all four *Paced RT* mean RT measures. Greater values of this factor indicated slower RTs on *Paced RT*. The second factor was referred to as "*Go No-Go Reaction Time Stability*" because it contained a high positive loading for *rRT* mean RT, as well as a very high positive and a very high negative loading for *rRT* coefficient of variation of RT and *rRT* standard deviation of RT, respectively. Greater values of this factor indicated more stable, slower RTs on *rRT*. The third factor, referred to as "*Paced RT Accuracy*," had high positive loadings for all four *Paced RT* accuracy measures. Greater values of this factor indicated greater accuracy on *Paced RT*. The fourth factor was referred to as "*Go No-Go Elite Performance*" because it contained very high positive loadings for both *rRT* stop rate and dprime, as well as a negative loading for *rRT* mean RT. High values of this factor indicated that the subject was capable of responding accurately to Go trials, inhibiting responses to No-Go trials, and responding quickly to Go trials. The fifth factor, referred to as "*Inhibition*," had all three *Rapid Adjust* measures (PSE, stop rate, and dprime) loading highly positively, which all measured the subject's ability to appropriately respond to Go trials, and inhibit responses on No-Go trials. Therefore, greater values of this factor indicated greater inhibitory control.

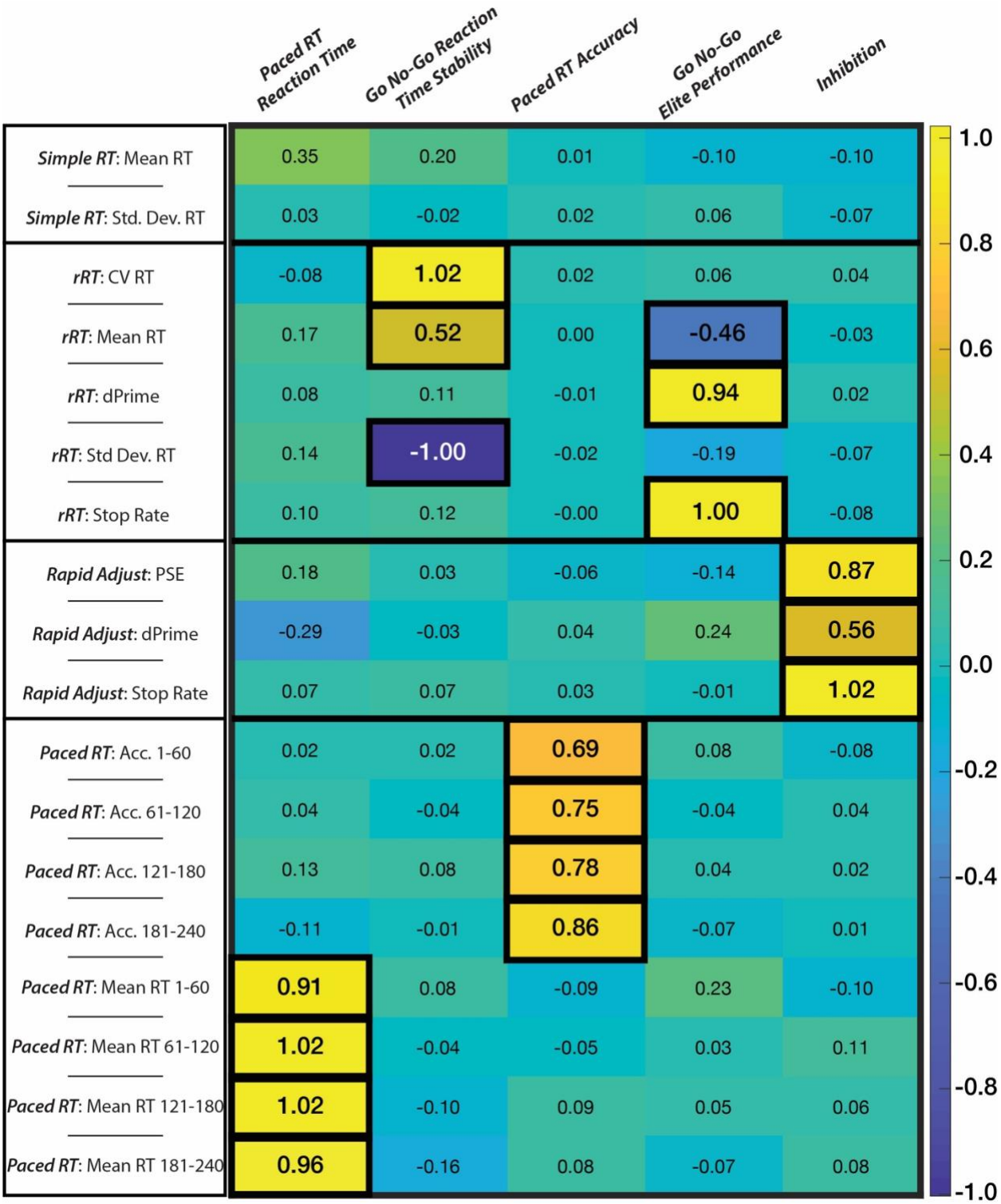


Figure 4.2 Factor loadings of behavioral measures. Shown are the factor loadings for each of the five derived factors across all 18 computed behavioral measures. Factors have been assigned names based on the behavioral measures that load most highly, with high loadings being highlighted to aid interpretability.

Relationship Between Questionnaire Responses and Behavior

Following completion of the four computerized tasks, subjects completed a questionnaire that assessed the manner in which these tasks were performed. For the *rRT* task, the questionnaire probed whether subjects prioritized speed or accuracy, or balanced both, while performing the task. For *Rapid Adjust*, the questionnaire asked if subjects experienced greater difficulty in responding to Go trials than inhibiting responses on No-Go trials. For *Paced RT*, the questionnaire asked if subjects felt they improved over the course of the task, whether they discovered the existence of the hidden repeating sequence, and their confidence in their discovery. We used these responses to assess if the manner that subjects approached the tasks affected their performance.

We first performed an ANCOVA using data from the *rRT* task. The model predicted values of Factor 2 ("*Go No-Go Reaction Time Stability"*) using responses to how subjects prioritized their performance (Priority: Accuracy, Speed, Balanced) as a fixed factor, controlling for values of reported alertness. The same approach was taken in predicting values of Factor 4 ("*Go No-Go Elite Performance"*). There was a main effect of Priority in predicting values of Factor 2 ($F(2,70) = 3.581$, $p = 0.033$), and a trending main effect in predicting values of Factor 4 ($F(2,70) = 2.691$, $p = 0.075$). Follow-up t-tests revealed that subjects who prioritized accuracy (N = 16, mean = 0.161, standard error = 0.243, $t(72) = 1.993$, $p = 0.05$), and those who took a balanced approach (N = 37, mean = 0.205, standard error = 0.16, $t(72) = 2.577$, $p = 0.012$) had greater values of Factor 2, when accounting for alertness, than those who prioritized speed (N = 21, mean = -0.483, standard error = 0.213). As performing follow-up t-tests would be statistically inappropriate for Factor 4 given the non-significant main effect of Priority, we instead performed an alternative analysis where we split the data into two groups based on whether or not a balanced approach was taken. Subjects who adopted a balanced approach (mean = 0.262, standard error = 0.161) had greater values of Factor 4 than subjects who did not (N = 37, mean = -0.272, standard error = 0.159, $t(72) = 2.421$, $p = 0.018$).

We next analyzed whether performance on *Paced RT* was affected by whether subjects thought they improved during the task (Improvement: No, Yes), whether they detected the hidden sequence (Sequence: No, Yes), and how confident they were in their detection of the sequence. We performed two ANCOVAs using values of Factor 1 ("*Paced RT Reaction Time"*) and Factor 3 ("*Paced RT Accuracy"*) as the dependent variables. The categorical responses to the two questionnaire items were entered as fixed factors, additionally modeling the interaction between the two, as well as accounting for the confidence in their response and their alertness. For both Factor 1 and Factor 3, we found no main effect of Improvement (Factor 1: $F(1,68) = 0.595$, $p = 0.443$); Factor 3: $F(1,68) = 1.488$, $p = 0.227$), Sequence (Factor 1: $F(1,68) = 0.899$, $p = 0.346$); Factor 3: $F(1,68) = 0.575$, $p = 0.451$), or the interaction of the two (Factor 1: $F(1,68) = 1.898$, $p = 0.173$); Factor 3: $F(1,68) = 0.43$, $p = 0.836$). Moreover, confidence was not a significant predictor of behavior either (Factor 1: $F(1,68) = 0.025$, $p = 0.874$); Factor 3: $F(1,68) = .146$, $p = 0.703$).

We did not analyze data for *Rapid Adjust* as there was insufficient variability in the questionnaire responses (88% of subjects responded that Go trial were more difficult).

Relationship Between Baseline Measures of Functional Connectivity and Physiology to Behavior

To assess how baseline measures of human functional brain organization track behavior, we computed global modularity, as well as the mean within-network functional connectivity for seven independent functional brain networks derived from the seven-network Yeo and colleagues parcellation (Visual, Somatomotor, Dorsal Attention, Ventral Attention / Salience, Limbic, Frontoparietal Control, and Default Mode; Yeo et al., 2011). In addition, we calculated various measures of baseline physiology, including the mean and standard deviation of HR and BR, as well as the RMSSD measure of HRV. We used these measures to predict behavioral performance using separate multiple regressions for each factor, while additionally accounting for values of reported alertness and mean FD (Tables 1-5).

We found that mean functional connectivity within the Frontoparietal Control network (β = -0.303 [-0.556, -0.050], p = 0.02), mean BR (β = 0.323 [0.092, 0.554], p = 0.007), and the standard deviation of BR (β = -0.301 [-0.541, -0.061], p = 0.015) predicted values of Factor 1 (Table 1). No baseline measures of functional brain organization or physiology predicted values of Factor 2 (Table 2). Measures of mean functional connectivity within the Motor network (β = 0.293 [0.064, 0.522], p = 0.013), as well as within the Ventral Attention / Salience network (β = -0.286 [-0.502, -0.007], p = 0.01) predicted values of Factor 3, while there existed a trend for the measure of modularity (β = 0.24 [-0.01, 0.49], p = 0.06; Table 3). There existed only a trending effect of mean connectivity within the Frontoparietal Control network (β = 0.239 [-0.045, 0.523], p = 0.098) in predicting values of Factor 4 (Table 4). Last, only the standard deviation of BR (β = 0.306 [0.052, 0.561], p = 0.019) predicted values of Factor 5 (Table 5).

Factor 1: ***"Paced RT Reaction Time"***	***β (95% CI)***	***p***
Modularity	0.079 [-0.184, 0.342]	0.551
Visual Netw. Conn.	0.118 [-0.133, 0.369]	0.351
Motor Netw. Conn.	0.121 [-0.12, 0.362]	0.319
DAN Netw. Conn.	-0.15 [-0.414, 0.113]	0.259
VAN / Sal. Netw. Conn.	0.055 [-0.172, 0.282]	0.628
Limbic Netw. Conn.	-0.144 [-0.404, 0.116]	0.272
FP Control Netw. Conn.	-0.303 [-0.556, -0.05]	0.02
DMN Netw. Conn.	0.068 [-0.222, 0.358]	0.641
Mean HR	0.114 [-0.181, 0.409]	0.443
Std. Dev. HR	0.174 [-0.082, 0.429]	0.179
Mean BR	0.323 [0.092, 0.554]	0.007
Std. Dev BR	-0.301 [-0.541, -0.061]	0.015
HRV	-0.091 [-0.374, 0.193]	0.525
Mean FD	-0.02 [-0.268, 0.227]	0.871
Alertness	-0.301 [-0.559, -0.043]	0.023

Table 4.1 Multiple regression of Factor 1 (*"Paced RT Reaction Time"*). Shown are standardized beta coefficients and their 95% confidence intervals, along with their significance values. Unique variance in both neural (FP Control network connectivity) and physiological (BR mean and variability) measures track values of Factor 1. DAN: Dorsal Attention network. DMN: Default Mode network. FP: Frontoparietal. VAN: Ventral Attention network. Yellow rows indicate $p < 0.05$.

Factor 2: ***"Go No-Go Reaction Time Stability"***	***β (95% CI)***	***p***
Modularity	0.096 [-0.212, 0.403]	0.535
Visual Netw. Conn.	0.099 [-0.195, 0.392]	0.504
Motor Netw. Conn.	0.078 [-0.203, 0.359]	0.582
DAN Netw. Conn.	-0.041 [-0.348, 0.267]	0.792
VAN / Sal. Netw. Conn.	0.04 [-0.225, 0.305]	0.765
Limbic Netw. Conn.	-0.064 [-0.368, 0.239]	0.672
FP Control Netw. Conn.	0.109 [-0.186, 0.404]	0.463
DMN Netw. Conn.	-0.009 [-0.348, 0.33]	0.959
Mean HR	-0.202 [-0.546, 0.143]	0.247
Std. Dev. HR	-0.192 [-0.49, 0.106]	0.202
Mean BR	-0.107 [-0.377, 0.162]	0.429
Std. Dev. BR	0.119 [-0.162, 0.399]	0.4
HRV	-0.112 [-0.443, 0.219]	0.501
Mean FD	-0.014 [-0.303, 0.275]	0.924
Alertness	0.039 [-0.262, 0.34]	0.796

Table 4.2 Multiple regression of Factor 2 (*"Go No-Go Reaction Time Stability"*). Shown are standardized beta coefficients and their 95% confidence intervals, along with their significance values. DAN: Dorsal Attention network. DMN: Default Mode network. FP: Frontoparietal. VAN: Ventral Attention network.

Factor 3: ***"Paced RT Accuracy"***	***β (95% CI)***	***p***
Modularity	0.24 [-0.01, 0.49]	0.06
Visual Netw. Conn.	-0.138 [-0.377, -0.101]	0.252
Motor Netw. Conn.	0.293 [0.064, 0.522]	0.013
DAN Netw. Conn.	0.144 [-0.106, 0.395]	0.253
VAN / Sal. Netw. Conn.	-0.286 [-0.502, -0.07]	0.01
Limbic Netw. Conn.	-0.045 [-0.293, 0.202]	0.715
FP Control Netw. Conn.	-0.017 [-0.257, 0.224]	0.89
DMN Netw. Conn.	-0.122 [-0.399, 0.154]	0.378
Mean HR	-0.049 [-0.33, 0.232]	0.729
Std. Dev. HR	-0.094 [-0.337, 0.148]	0.439
Mean BR	0.074 [-0.145, 0.294]	0.5
Std. Dev. BR	0.131 [-0.097, 0.36]	0.254
HRV	-0.046 [-0.316, 0.223]	0.733
Mean FD	-0.233 [-0.469, 0.002]	0.052
Alertness	-0.25 [-0.496, -0.005]	0.046

Table 4.3 Multiple regression of Factor 3 (*"Paced RT Accuracy"*). Shown are standardized beta coefficients and their 95% confidence intervals, along with their significance values. Unique variance in measures of both global (modularity) and meso-scale (network connectivites) brain organization predict values of Factor 3. DAN: Dorsal Attention network. DMN: Default Mode network. FP: Frontoparietal. VAN: Ventral Attention network. Yellow rows indicate $p < 0.05$, blue indicates $p < 0.10$.

Factor 4: ***"Go No-Go Elite Performance"***	***β (95% CI)***	***p***
Modularity	-0.147 [-0.443, 0.148]	0.323
Visual Netw. Conn.	-0.037 [-0.319, 0.245]	0.794
Motor Netw. Conn.	-0.008 [-0.278, 0.262]	0.953
DAN Netw. Conn.	-0.062 [-0.358, 0.233]	0.674
VAN / Sal. Netw. Conn.	-0.09 [-0.345, 0.165]	0.483
Limbic Netw. Conn.	-0.046 [-0.338, 0.246]	0.754
FP Control Netw. Conn.	0.239 [-0.045, 0.523]	0.098
DMN Netw. Conn.	-0.108 [-0.434, 0.218]	0.51
Mean HR	-0.184 [-0.515, 0.148]	0.272
Std. Dev. HR	-0.151 [-0.437, 0.135]	0.296
Mean BR	-0.19 [-0.449, 0.069]	0.148
Std. Dev. BR	0.129 [-0.14, 0.399]	0.34
HRV	-0.048 [-0.367, 0.27]	0.762
Mean FD	-0.019 [-0.297, 0.259]	0.891
Alertness	0.174 [-0.115, 0.464]	0.233

Table 4.4 Multiple regression of Factor 4 (*"Go No-Go Elite Performance"*). Shown are standardized beta coefficients and their 95% confidence intervals, along with their significance values. A trend may exist between unique variance in mean functional connectivity within the Frontoparietal Control network and values of Factor 4. DAN: Dorsal Attention network. DMN: Default Mode network. FP: Frontoparietal. VAN: Ventral Attention network. Blue rows indicate $p < 0.10$.

Factor 5: ***"Inhibition"***	***β (95% CI)***	***p***
Modularity	-0.052 [-0.331 – 0.227]	0.711
Visual Netw. Conn.	-0.193 [-0.459, 0.074]	0.153
Motor Netw. Conn.	-0.009 [-0.264, 0.247]	0.946
DAN Netw. Conn.	0.23 [-0.049, 0.51]	0.104
VAN / Sal. Netw. Conn.	-0.114 [-0.355, 0.127]	0.347
Limbic Netw. Conn.	0.061 [-0.215, 0.337]	0.661
FP Control Netw. Conn.	0.122 [-0.147, 0.39]	0.367
DMN Netw. Conn.	-0.061 [-0.37, 0.247]	0.691
Mean HR	-0.25 [-0.563, 0.064]	0.116
Std. Dev. HR	-0.16 [-0.431, 0.11]	0.24
Mean BR	-0.2 [-0.445, 0.045]	0.107
Std. Dev. BR	0.306 [0.052, 0.561]	0.019
HRV	-0.003 [-0.304, 0.298]	0.984
Mean FD	-0.015 [-0.277, 0.248]	0.912
Alertness	0.293 [0.019, 0.567]	0.036

Table 4.5 Multiple regression of Factor 5 ("*Inhibition"*). Shown are standardized beta coefficients and their 95% confidence intervals, along with their significance values. Unique variance in the variability of BR is revealed to positively predict values of Factor 5. DAN: Dorsal Attention network. DMN: Default Mode network. FP: Frontoparietal. VAN: Ventral Attention network. Yellow rows indicate $p < 0.05$.

Task Approach Associated with Unique Brain- and Physio-Behavior Relationships for rRT

Based on the behavioral differences observed in the *rRT* task behavior (Factors 2 and 4) when considering subjects' responses to the post-test questionnaire, we assessed whether there existed unique relationships between behavior and biology within each grouping. We split the data by responses to the Priority question and assessed the Pearson correlation between behavior (Factors 2 and 4) and our measures of baseline functional brain organization and physiology. For Factor 2, we found no significant relationships for either the accuracy subgroup or speed subgroup. However, we found a negative relationship between mean HR (r = -0.331, p = 0.045, Figure 3A) and values of Factor 2 in the subgroup of subjects who adopted a balanced approach to the *rRT* task.

For Factor 4 values, we found that mean functional connectivity within the Ventral Attention / Salience network (r = -0.623, p = 0.010, Figure 3B) negatively predicted behavior in the accuracy subgroup. However, in the speed subgroup, there existed a positive relationship with mean functional connectivity within the Ventral Attention / Salience network (r = 0.582, p = 0.006, Figure 3C). In addition, there were trending relationships between Factor 4 values and both modularity (r = -0.382, p = 0.087, Figure 3D) and mean functional connectivity within the Frontoparietal Control network (r = 0.384, p = 0.086, Figure 3E) in the speed subgroup. Last, we again found a negative relationship with mean HR (r = -0.338, p = 0.041, Figure 3F) for subjects who adopted a balanced approach.

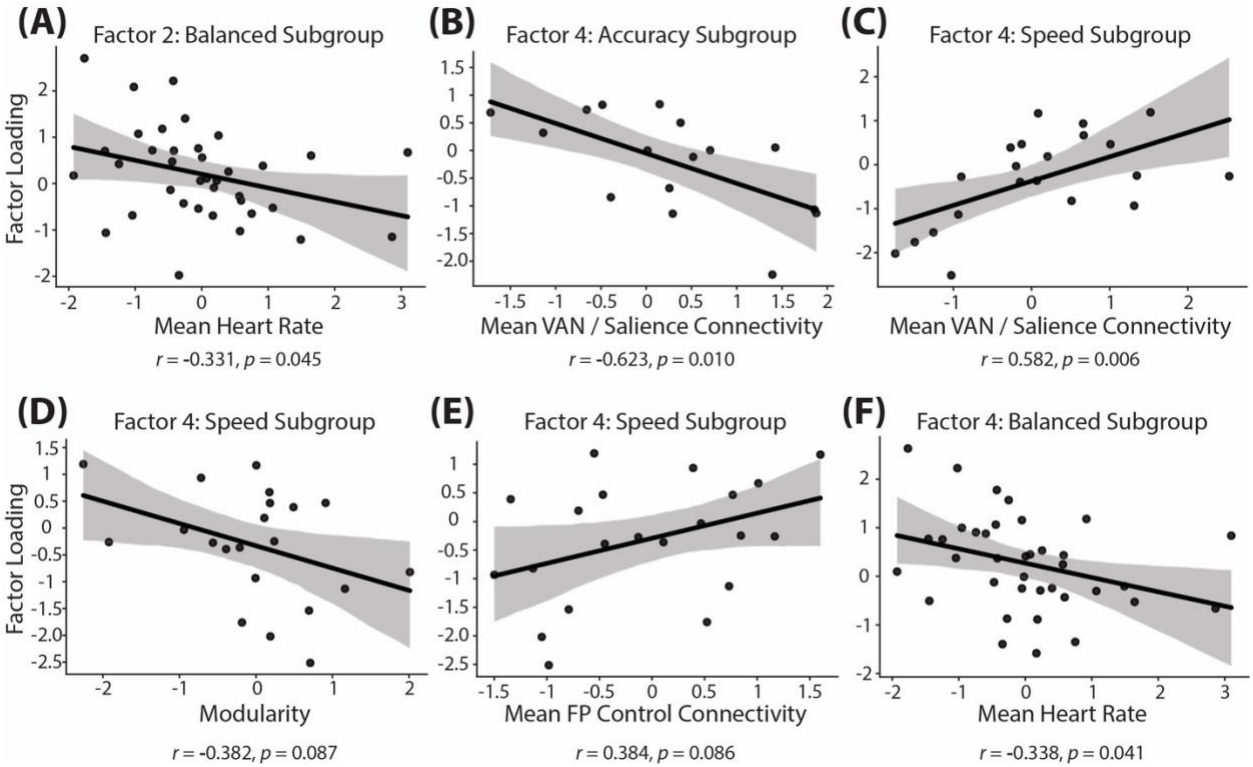


Figure 4.3 Scatterplots of Relationships within separate *Paced RT* subgroups. (A) Factor 2 is negatively predicted by mean HR in the balanced subgroup (N = 37). (B, C) Mean functional connectivity within the Ventral Attention / Salience network in the (B) accuracy subgroup (N = 16) negatively predicts Factor 4, and (C) positively predicts Factor 4 in the speed subgroup (N = 31). (D) Trending negative relationship between Factor 4 and modularity in the speed subgroup. (E) Trending positive relationship between mean functional connectivity within the Frontoparietal Control network and Factor 4 in the speed subgroup. (F) Similar to Factor 2, mean HR also negatively predicts Factor 4 in the balance subgroup. All values shown are standardized units.

4.5 Discussion

The current report collected a large dataset spanning behavior, physiology, and neural activity in order to assess the ability of baseline biological measures to track cognitive abilities required for successful baseball performance. We found that basic RT was predicted by both physiological measures and a network-level neural measure, while the behavioral factor related to RT variability was tracked only by a single physiological measure. Basic accuracy was predicted only by neural measures, at both the brain-wide and network-level. Inhibitory control, regardless of RT, was tracked by a single physiological measure, while a behavioral factor representing both inhibitory control and RT showed unique relationships depending upon the strategy adopted by the subject. These results show that neither neural nor physiological measures alone are sufficient to predict all behavioral abilities central to baseball performance.

Successfully hitting a baseball requires more than just quick RT. Within milliseconds the batter must make a judgment on the type of pitch being thrown, the upcoming location of the ball, and whether or not to swing. While rapid RT is indeed a necessary skill, the batter must also be able to discriminate between pitch types given highly correlated visual input: all things

considered, the kinetics of a professional pitcher releasing a fastball are very similar to that of a change-up. Thus, speeded RT must be paired with accurate decision making if one is to successfully swing at the ball (i.e. assessed with *rRT*). However, given the trajectory of the ball, the batter may decide instead to withhold their swing due to the ball falling outside of the strike zone (i.e., assessed with *Rapid Adjust*). These scenarios require that batters rapidly inhibit their prepotent response to swing. All of these additionally rely on the known repertoire of pitches available to the pitcher and the relative likelihood of each pitch being thrown in the current context (i.e., assessed with *Paced RT*). The behavioral tasks employed in the current study tap into all of these processes. Moreover, the factor analysis resulted in an intuitive clustering of behavior within each task.

The first factor contained the RT measures from the *Paced RT* task, largely representing the baseline speed with which one responds to a visual probe. Values of this factor did not vary according to whether the subjects felt they improved on the task over time, or on their subjective belief in the presence of a repeating pattern across the trials. The average functional connectivity within the Frontoparietal Control network negatively tracked values of this factor, indicating that greater connectivity within this network predicted faster response times. This functional network has been associated with moment-to-moment processing of task-relevant information (Dixon et al., 2018; Dosenbach et al., 2008), and as such, greater connectivity within this network at baseline may indicate a greater ability to rapidly process incoming information and enact appropriate motor commands. Values of this factor were also tracked by the mean and standard deviation of the BR. A greater mean BR was associated with greater (slower) RTs, while greater variability in BR was associated with faster RTs. These results are somewhat difficult to interpret considering that respiratory control can be under conscious and unconscious control during the baseline measurement period. Since measures of physiology were acquired during the resting-state fMRI scans, these values may better represent how the subject is able to cope with unfamiliar situations, such as being inside the bore of the magnet.

The second factor contained a highly negative loading for the measure of the standard deviation of RT, as well as a relatively high positive loading for mean RT, in the *rRT* task. Therefore, high values of this factor represent the intersection of stable, and to a lesser extent, slow RTs on correct response trials for a task of inhibitory control. Unsurprisingly, subjects who prioritized speed while performing the task had the lowest values of this factor compared to those prioritizing either accuracy or a balance between speed and accuracy. Across all subjects, neither neural nor physiological measures tracked values of this factor. However, in separating the data into subgroups based on task priority, we found that those who adopted a balanced approach showed a negative relationship between mean HR and the factor values. In other words, a lower mean HR at baseline was associated with more stable, slowed correct responses in those who adopted a balanced response approach. This result is intuitive as greater control over one's physiological state is likely to align with greater stability when responding rapidly to incoming information.

The third factor contained positive loadings for the accuracy measures from the *Paced RT* task, and thus represents a basic form of accuracy in a task where both speed and accuracy of responses are valued. This factor was only predicted by measures of neural activity. Greater functional connectivity within the Motor network was associated with greater accuracy across

the trials of the *Paced RT* task. As this task required the execution of fast action mappings between the on-screen stimulus and one of four button responses, greater functional connectivity within this network likely aligns with a greater baseline ability to enact the correct motor action in a timely manner. In addition, greater functional connectivity within the Ventral Attention / Salience network was associated with lower accuracy in this task. Activity throughout this network is associated with alerting or reorienting of attention to salient stimuli in the environment (Kastner, 2017). Higher functional connectivity throughout this network while at rest may indicate a greater rate of reorienting of one's attention to stimuli deemed relevant in the current context. As such, this reorienting could suggest a lesser ability to sustain attention and therefore may predict worse performance on a task that requires minimal amounts of attentional shifting. Last, there was a trending positive association between brain-wide modularity and values of Factor 3. Modularity has been previously associated with the ability to learn new information and is predictive of cognitive training improvement (Bassett & Mattar, 2017; Gallen & Esposito, 2019). The current results align with these findings as high accuracy is largely dependent on learning the correct task-specific visuomotor response mappings.

The fourth factor contained high loadings for a mix of accuracy and response time measures for the *rRT* task. High values of this factor represented greater inhibition on No-Go trials as well as both greater signal detection and more rapid responses on Go trials. Overall, high values of this factor indicated greater performance on the *rRT* task in all facets. Across all subjects, there existed a trending positive association between values of this factor and functional connectivity within the Frontoparietal Control network. However, when separating subjects based on how they approached the task, we discovered unique relationships for each group. There existed a negative relationship between mean functional connectivity within the Ventral Attention / Salience network and values of this factor for subjects prioritizing accuracy. However, the opposite existed for subjects who prioritized speed: greater levels of mean functional connectivity within the Ventral Attention / Salience network at rest predicted greater values of this factor. The Salience network is associated with the coordination of task-relevant behavioral responses as well as to acute stress reactivity (Hermans, Henckens, Joëls, & Fernández, 2014; Seeley, 2019; Seeley et al., 2007). Subjects who prioritized different task-relevant components during task performance are potentially likely to benefit from baseline, trait-level Salience network connectivity in unique ways, and as such unique, and potentially opposite, relationships with behavioral performance may exist. The speed subgroup also showed a trending negative and positive relationship for modularity and mean functional connectivity within the Frontoparietal Control network, respectively. Last, and mirroring the relationship seen with Factor 2, mean resting HR negatively predicted values of Factor 4 in the subgroup of subjects who adopted a balanced approach. It is therefore evident that lower levels of baseline HR relate to greater task performance when subjects equally prioritize task-relevant features. Overall, a trait disposition to unequally prioritize task-relevant features may result in a lesser ability to maximize task performance, evidenced by greater performance in the balanced group compared to the others. However, individual differences in performance for low-performing subgroups are still uniquely related to baseline differences in functional brain organization.

The fifth factor contained high positive loadings for all three behavioral measures from the *Rapid Adjust* task, and specifically contained the highest loadings for the two measures that tracked inhibitory control (i.e., PSE and Stop Rate). Values of this factor were only associated

with a single measure of physiological functioning: greater variability in one's resting BR predicted greater values of Factor 5. Similar to the association with Factor 1 values, greater variability in one's resting BR is again associated with greater task performance. It may be reasonable to assume that greater variability in BR is the result of a greater degree of switching between states of conscious and unconscious control of one's respiration during the measurement period, as breathing patterns and rates are likely to be quite different between the two states. Potentially, the balance of switching between these states may be a trait-level marker of physiological functioning that indexes something unique about cognitive abilities. Further investigations on this topic will be necessary to better elucidate these results.

In conclusion, the current study presents evidence that although cognitive abilities crucial for successful baseball performance are predicted by measures of baseline physiology and functional brain organization, neither measurement modality is sufficient to track all necessary abilities. Neural measures reflecting network-specific properties tracked basic RT and accuracy, as well as elite inhibitory performance in certain subgroups. Neural measures reflecting brain-wide functional organization also tracked basic accuracy and subgroup-specific elite inhibitory performance. Although tracking a behavioral factor that incorporates both inhibition accuracy and RT, no neural measures tracked inhibition accuracy or inhibition RT separately. However, measures of baseline physiological functioning were able to track these behaviors separately, as well as basic RT and subgroup-specific elite inhibitory performance. Therefore, in order to achieve a comprehensive assessment of an individual's baseball-related skills, it is necessary to collect both measures of physiology and functional neural organization.

CHAPTER 5

5.1 Conclusions and Future Directions

Across these three experiments I employed multiple biological analysis methods and behavioral models in order to investigate the neural underpinnings associated with multiple complex human behaviors. These methods included the analysis of fMRI data while subjects performed cognitive tasks, as well as when subjects were at rest. Moreover, behavioral data from numerous cognitive tasks was analyzed in a unique manner for each experiment. The variability in these approaches allowed for the interrogation of crucial brain-behavioral relationships.

In Chapter 2, I used task-based fMRI and computational modeling of task behavior to determine which brain regions support the discovery and subsequent generalization of a hidden hierarchically-structured task rule. I found that multiple brain regions spanning frontal and parietal cortices supported the initial discovery of the hidden rule, while regions spanning a cingulo-opercular network tracked individual differences in the ability to generalize this rule to novel contexts. These results provide original evidence for each network's role in achieving the coordinated set of behaviors necessary to extract relevant information from one's environment and to utilize this knowledge in the future. It is necessary to further explore the manner in which these networks, and potentially others, interact with one another to achieve these behaviors. Moreover, causal methods, such as transcranial magnetic stimulation (TMS), should be employed in order to draw a causal link between these networks and their respective functions.

In Chapters 3 and 4, I analyzed fMRI data collected during a task-free resting-state in order to build predictive models of complex human behavior. In Chapter 3 I investigated the differences between traditional time-invariant and novel time-varying analysis approaches. As predicted, time-varying analyses better captured individual differences in measures of behavioral variability, while time-invariant measures tracked more stable, task-general cognitive abilities. In Chapter 4, I compared the predictive power of neural measures against measures indexing physiological functioning. It was revealed that neural measures tracked behaviors related to basic accuracy and reaction time, while physiological measures were better able to track accuracy and reaction time of inhibition-based behaviors. Further analysis of fMRI data within this context would prove fruitful. First, expanding these approaches to datasets of task-based fMRI would allow for task manipulations to inform unique temporally-varying patterns of functional connectivity. Second, investigating whether these results are seen in clinical populations, such as those with neurological pathology or psychiatric diagnoses, would allow for a more refined understanding of the neural activity patterns underlying those conditions.

References

Abdelmounaime, S., & Dong-Chen, H. (2013). New Brodatz-Based Image Databases for Grayscale Color and Multiband Texture Analysis. *ISRN Machine Vision*, *2013*, 1–14. https://doi.org/10.1155/2013/876386

Alexander-Bloch, A. F., Gogtay, N., Meunier, D., Birn, R., Clasen, L., Lalonde, F., … Bullmore, E. T. (2010). Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. *Frontiers in Systems Neuroscience*, *4*(October), 1–16. https://doi.org/10.3389/fnsys.2010.00147

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience*, *19*, 2082–2099. https://doi.org/10.1162/jocn.2007.19.12.2082

Badre, D., & Frank, M. J. (2012). Mechanisms of hierarchical reinforcement learning in cortico-striatal circuits 2: Evidence from fMRI. *Cerebral Cortex*, *22*(3), 527–536. https://doi.org/10.1093/cercor/bhr117

Badre, D., Hoffman, J., Cooney, J. W., & D'Esposito, M. (2009). Hierarchical cognitive control deficits following damage to the human frontal lobe. *Nature Neuroscience*, *12*(4), 515–522. https://doi.org/10.1038/nn.2277

Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal Cortex and the Discovery of Abstract Action Rules. *Neuron*, *66*(2), 315–326. https://doi.org/10.1016/j.neuron.2010.03.025

Badre, D., & Nee, D. E. (2018). Frontal Cortex and the Hierarchical Control of Behavior. *Trends in Cognitive Sciences*, *22*(2), 170–188. https://doi.org/10.1016/j.tics.2017.11.005

Baker, A. P., Brookes, M. J., Rezek, I. A., Smith, S. M., Behrens, T., Smith, P. J. P., & Woolrich, M. (2014). Fast transient networks in spontaneous human brain activity. *ELife*, *2014*(3), 1–18. https://doi.org/10.7554/eLife.01867

Bassett, D. S., & Mattar, M. G. (2017). A Network Neuroscience of Human Learning: Potential to Inform Quantitative Theories of Brain and Behavior. *Trends in Cognitive Sciences*, *21*(4), 250–264. https://doi.org/10.1016/j.tics.2017.01.010

Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: learning to learn and action video games. *Annual Reviews of Neuroscience*, *35*, 391–416. https://doi.org/10.1146/annurev-neuro-060909

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Biswal, B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, *34*(4), 537–541. https://doi.org/10.1002/mrm.1910340409

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.

https://doi.org/10.1037/0033-295X.108.3.624
Botvinick, M., Ritter, S., Wang, J. X., Kurth-nelson, Z., Blundell, C., & Hassabis, D. (2019). Reinforcement Learning , Fast and Slow. *Trends in Cognitive Sciences*, 1–15. https://doi.org/10.1016/j.tics.2019.02.006
Braga, R. M., Sharp, D. J., Leeson, C., Wise, R. J. S., & Leech, R. (2013). Echoes of the Brain within Default Mode, Association, and Heteromodal Cortices. *Journal of Neuroscience*, *33*(35), 14031–14039. https://doi.org/10.1523/JNEUROSCI.0570-13.2013
Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox. In *8th International Conference on Functional Mapping of the Human Brain*.
Brookes, M. J., O'Neill, G. C., Hall, E. L., Woolrich, M. W., Baker, A., Palazzo Corner, S., … Barnes, G. R. (2014). Measuring temporal, spectral and spatial changes in electrophysiological brain network connectivity. *NeuroImage*, *91*, 282–299. https://doi.org/10.1016/j.neuroimage.2013.12.066
Burris, K., Vittetoe, K., Ramger, B., Suresh, S., Tokdar, S. T., Reiter, J. P., & Appelbaum, L. G. (2018). Sensorimotor abilities predict on-field performance in professional baseball. *Scientific Reports*, *8*(1), 116. https://doi.org/10.1038/s41598-017-18565-7
Calhoun, V. D., Miller, R., Pearlson, G., & Adali, T. (2014). The Chronnectome: Time-Varying Connectivity Networks as the Next Frontier in fMRI Data Discovery. *Neuron*, *84*(2), 262–274. https://doi.org/10.1016/j.neuron.2014.10.015
Casorso, J., Kong, X., Chi, W., Van De Ville, D., Yeo, B. T. T., & Liégeois, R. (2019). Dynamic mode decomposition of resting-state and task fMRI. *NeuroImage*, *194*(February), 42–54. https://doi.org/10.1016/j.neuroimage.2019.03.019
Chan, M. Y., Park, D. C., Savalia, N. K., Petersen, S. E., & Wig, G. S. (2014). Decreased segregation of brain systems across the healthy adult lifespan. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(46), E4997-5006. https://doi.org/10.1073/pnas.1415122111
Chang, C., & Glover, G. H. (2010). Time–frequency dynamics of resting-state brain connectivity measured with fMRI. *NeuroImage*, *50*(1), 81–98. https://doi.org/10.1016/j.neuroimage.2009.12.011
Choi, E. Y., Drayna, G. K., & Badre, D. (2018). Evidence for a Functional Hierarchy of Association Networks. *Journal of Cognitive Neuroscience*, *30*(5), 722–736. https://doi.org/10.1162/jocn_a_01229
Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., … Satterthwaite, T. D. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*, *154*(March), 174–187. https://doi.org/10.1016/j.neuroimage.2017.03.020
Cisler, J. M., James, G. A., Tripathi, S., Mletzko, T., Heim, C., Hu, X. P., … Kilts, C. D. (2013). Differential functional connectivity within an emotion regulation neural network among individuals resilient and susceptible to the depressogenic effects of early life stress. *Psychological Medicine*, *43*(3), 507–518. https://doi.org/10.1017/S0033291712001390
Cohen, J. R. (2018). The behavioral and cognitive relevance of time-varying, dynamic changes in functional connectivity. *NeuroImage*, *180*(April 2017), 515–525. https://doi.org/10.1016/j.neuroimage.2017.09.036
Cohen, J. R., & D'Esposito, M. (2016). The Segregation and Integration of Distinct Brain Networks and Their Relationship to Cognition. *Journal of Neuroscience*, *36*(48), 12083–

12094. https://doi.org/10.1523/JNEUROSCI.2965-15.2016
Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., & Petersen, S. E. (2014). Intrinsic and task-evoked network architectures of the human brain. *Neuron*, *83*(1), 238–251. https://doi.org/10.1016/j.neuron.2014.05.014
Collins, A. G. E., Cavanagh, J. F., & Frank, M. J. (2014). Human EEG Uncovers Latent Generalizable Rule Structure during Learning. *The Journal of Neuroscience*, *34*(13), 4677–4685. https://doi.org/10.1523/JNEUROSCI.3900-13.2014
Collins, A. G. E., & Frank, M. J. (2013). Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychological Review*, *120*(1), 190–229. https://doi.org/10.1037/a0030852
Collins, A. G. E., & Frank, M. J. (2016). Neural signature of hierarchically structured expectations predicts clustering and transfer of rule sets in reinforcement learning. *Cognition*, *152*, 160–169. https://doi.org/10.1016/j.cognition.2016.04.002
Cornell, D. J., Paxson, J. L., Caplinger, R. A., Seligman, J. R., Davis, N. A., & Ebersole, K. T. (2017). Resting Heart Rate Variability Among Professional Baseball Starting Pitchers. *Journal of Strength and Conditioning Research*, *31*(3), 575–581. https://doi.org/10.1519/JSC.0000000000001538
Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, *10*(4–5), 171–178. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L
D'Esposito, M., & Postle, B. R. (2015). The Cognitive Neuroscience of Working Memory. *Annual Review of Psychology*, *66*(1), 115–142. https://doi.org/10.1146/annurev-psych-010814-015031
Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, *8*(2–3), 109–114. https://doi.org/10.1002/(SICI)1097-0193(1999)8:2/3<109::AID-HBM7>3.0.CO;2-W
Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395
de Pasquale, F., Della Penna, S., Snyder, A. Z., Lewis, C., Mantini, D., Marzetti, L., … Corbetta, M. (2010). Temporal dynamics of spontaneous MEG activity in brain networks. *Proceedings of the National Academy of Sciences*, *107*(13), 6040–6045. https://doi.org/10.1073/pnas.0913863107
Dixon, M. L., De La Vega, A., Mills, C., Andrews-Hanna, J., Spreng, R. N., Cole, M. W., & Christoff, K. (2018). Heterogeneity within the frontoparietal control network and its relationship to the default and dorsal attention networks. *Proceedings of the National Academy of Sciences*, *115*(7), E1598–E1607. https://doi.org/10.1073/pnas.1715766115
Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, *12*(3), 99–105. https://doi.org/10.1016/j.tics.2008.01.001
Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., … Petersen, S. E. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences*, *104*(26), 11073–11078. https://doi.org/10.1073/pnas.0704320104
Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., … Petersen, S. E. (2006). A Core System for the Implementation of Task Sets. *Neuron*, *50*(5), 799–812. https://doi.org/10.1016/j.neuron.2006.04.031

Duff, E. P., Makin, T., Cottaar, M., Smith, S. M., & Woolrich, M. W. (2018). Disambiguating brain functional connectivity. *NeuroImage*, *173*, 540–550. https://doi.org/10.1016/j.neuroimage.2018.01.053

Duran, G., Tapiero, I., & Michael, G. A. (2018). Resting heart rate: A physiological predicator of lie detection ability. *Physiology & Behavior*, *186*, 10–15. https://doi.org/10.1016/j.physbeh.2018.01.002

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., … Gorgolewski, K. J. (2018). poldracklab/fmriprep: 1.0.10. https://doi.org/10.5281/ZENODO.1219187

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., … Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47*, S102. https://doi.org/10.1016/S1053-8119(09)70884-5

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). From The Cover: The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, *102*(27), 9673–9678. https://doi.org/10.1073/pnas.0504136102

Frank, M. J., & Badre, D. (2012). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: Computational analysis. *Cerebral Cortex*, *22*(3), 509–526. https://doi.org/10.1093/cercor/bhr114

Fuster, J. M. (1973). Unit activity in prefrontal cortex during delayed-response performance: Neuronal correlates of transient memory, 61–78.

Gallen, C. L., & Esposito, M. D. (2019). Brain Modularity : A Biomarker of Intervention-related Plasticity. *Trends in Cognitive Sciences*, *23*(4), 293–304. https://doi.org/10.1016/j.tics.2019.01.014

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Frontiers in Neuroinformatics*, *5*(August). https://doi.org/10.3389/fninf.2011.00013

Grassmann, M., Vlemincx, E., von Leupoldt, A., Mittelstädt, J. M., & Van den Bergh, O. (2016). Respiratory Changes in Response to Cognitive Load: A Systematic Review. *Neural Plasticity*, *2016*, 1–16. https://doi.org/10.1155/2016/8146809

Greve, D. N., & Fischl, B. (2009a). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Greve, D. N., & Fischl, B. (2009b). NeuroImage Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Hansen, A. L., Johnsen, B. H., & Thayer, J. F. (2003). Vagal influence on working memory and attention. *International Journal of Psychophysiology*, *48*(3), 263–274. https://doi.org/10.1016/S0167-8760(03)00073-4

Harlow, H. F. (1949). The formation of learning sets. *Psychological Review*, *56*(1), 51–65. https://doi.org/10.1037/h0062474

Hastie, T., Friedman, J. H. ., & Tibshirani, R. (2009). *The Elements of Statistical Learning*. New York City, New York: Springer US.

Henson, R. (2006). Forward inference using functional neuroimaging: dissociations versus associations. *Trends in Cognitive Sciences*, *10*(2), 64–69. https://doi.org/10.1016/j.tics.2005.12.005

Hermans, E. J., Henckens, M. J. A. G., Joëls, M., & Fernández, G. (2014). Dynamic adaptation of large-scale brain networks in response to acute stressors. *Trends in Neurosciences*, *37*(6), 304–314. https://doi.org/10.1016/j.tins.2014.03.006

Hindriks, R., Adhikari, M. H., Murayama, Y., Ganzetti, M., Mantini, D., Logothetis, N. K., & Deco, G. (2016). Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI? *NeuroImage*, *127*, 242–256. https://doi.org/10.1016/j.neuroimage.2015.11.055

Hotelling, H. (1936). Relations Between Two Sets of Variates. *Biometrika*, *28*(3–4), 321–377. https://doi.org/10.1093/biomet/28.3-4.321

Hovland, A., Pallesen, S., Hammar, Å., Hansen, A. L., Thayer, J. F., Tarvainen, M. P., & Nordhus, I. H. (2012). The relationships among heart rate variability, executive functions, and clinical variables in patients with panic disorder. *International Journal of Psychophysiology*, *86*(3), 269–275. https://doi.org/10.1016/j.ijpsycho.2012.10.004

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., … Chang, C. (2013). Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage*, *80*(1–2), 360–378. https://doi.org/10.1016/j.neuroimage.2013.05.079

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002a). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1016/S1053-8119(02)91132-8

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002b). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images, *841*, 825–841. https://doi.org/10.1006/nimg.2002.1132

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6

Jia, H., Hu, X., & Deshpande, G. (2014). Behavioral relevance of the dynamics of the functional brain connectome. *Brain Connectivity*, *4*(9), 741–759. https://doi.org/10.1089/brain.2014.0300

Kastner, S. (2017). Evolution of Visual Attention in Primates. In *Evolution of Nervous Systems* (pp. 237–256). Elsevier. https://doi.org/10.1016/B978-0-12-804042-3.00083-X

Kayser, A. S., & D'Esposito, M. (2013). Abstract rule learning: The differential effects of lesions in frontal cortex. *Cerebral Cortex*, *23*(1), 230–240. https://doi.org/10.1093/cercor/bhs013

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2010). Learning to learn causal models. *Cognitive Science*, *34*(7), 1185–1243. https://doi.org/10.1111/j.1551-6709.2010.01128.x

Kennedy, D. O., & Scholey, A. B. (2000). Glucose administration, heart rate and cognitive performance: effects of increasing mental effort. *Psychopharmacology*, *149*(1), 63–71. https://doi.org/10.1007/s002139900335

Khambhati, A. N., Medaglia, J. D., Karuza, E. A., Thompson-Schill, S. L., & Bassett, D. S. (2018). *Subgraphs of functional brain networks identify dynamical constraints of cognitive control*. *PLoS Computational Biology* (Vol. 14). https://doi.org/10.1371/journal.pcbi.1006234

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., … Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2),

e1005350. https://doi.org/10.1371/journal.pcbi.1005350
Koechlin, E. (2003). The Architecture of Cognitive Control in the Human Prefrontal Cortex. *Science*, *302*(5648), 1181–1185. https://doi.org/10.1126/science.1088545
Kucyi, A., Tambini, A., Sadaghiani, S., Keilholz, S., & Cohen, J. R. (2018). Spontaneous cognitive processes and the behavioral validation of time-varying brain connectivity. *Network Neuroscience*, *2*(4), 397–417. https://doi.org/10.1162/netn_a_00037
Kuo, B.-C., Yeh, Y.-Y., Chen, A. J.-W., & D'Esposito, M. (2011). Functional connectivity during top-down modulation of visual short-term memory representations. *Neuropsychologia*, *49*(6), 1589–1596. https://doi.org/10.1016/j.neuropsychologia.2010.12.043
Lanczos, C. (1964). Evaluation of Noisy Data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, *1*(1), 76–85. https://doi.org/10.1137/0701007
Liegeois, R., Li, J., Kong, R., Orban, C., Van De Ville, D., Ge, T., … Yeo, B. T. T. (2019). Resting brain dynamics at different timescales capture distinct aspects of human behavior. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-10317-7
Lindquist, M. A., Xu, Y., Nebel, M. B., & Caffo, B. S. (2014). Evaluating dynamic bivariate correlations in resting-state fMRI: A comparison study and a new approach. *NeuroImage*, *101*, 531–546. https://doi.org/10.1016/j.neuroimage.2014.06.052
Lurie, D. J., Kessler, D., Bassett, D. S., Betzel, R. F., Breakspear, M., Keilholz, S., … Calhoun, V. D. (2019). Questions and controversies in the study of time-varying functional connectivity in resting fMRI. *Network Neuroscience*, 1–40. https://doi.org/10.1162/netn_a_00116
Makris, N., Goldstein, J. M., Kennedy, D., Hodge, S. M., Caviness, V. S., Faraone, S. V., … Seidman, L. J. (2006). Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophrenia Research*, *83*(2–3), 155–171. https://doi.org/10.1016/j.schres.2005.11.020
Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, *19*(3), 1233–1239. https://doi.org/10.1016/S1053-8119(03)00169-1
McCormack, H. M., Horne, D. J., & Sheather, S. (1988). Clinical applications of visual analogue scales: a critical review. *Psychological Medicine*, *18*(4), 1007–1019. https://doi.org/10.1017/s0033291700009934
Muraskin, J., Dodhia, S., Lieberman, G., Garcia, J. O., Verstynen, T., Vettel, J. M., … Sajda, P. (2016). Brain dynamics of post-task resting state are influenced by expertise: Insights from baseball players. *Human Brain Mapping*, *37*(12), 4454–4471. https://doi.org/10.1002/hbm.23321
Nee, D. E., & D'Esposito, M. (2016). The hierarchical organization of the lateral prefrontal cortex. *ELife*, *5*(MARCH2016), 1–26. https://doi.org/10.7554/eLife.12112
Park, G., & Thayer, J. F. (2014). From the heart to the mind: cardiac vagal tone modulates top-down and bottom-up visual perception and attention to emotional stimuli. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00278
Park, G., Vasey, M. W., Van Bavel, J. J., & Thayer, J. F. (2013). Cardiac vagal tone is correlated with selective attention to neutral distractors under load. *Psychophysiology*, *50*(4), 398–406. https://doi.org/10.1111/psyp.12029
Parkes, L., Fulcher, B., Yücel, M., & Fornito, A. (2018). An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI.

NeuroImage, *171*(December 2017), 415–436. https://doi.org/10.1016/j.neuroimage.2017.12.073

Patanaik, A., Tandi, J., Ong, J. L., Wang, C., Zhou, J., & Chee, M. W. L. (2018). Dynamic functional connectivity and its behavioral correlates beyond vigilance. *NeuroImage*, *177*, 1–10. https://doi.org/10.1016/j.neuroimage.2018.04.049

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology*, *51*(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:6<768::aid-jclp2270510607>3.0.co;2-1

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*(January), 1–8. https://doi.org/10.3389/neuro.11.010.2008

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, *59*(3), 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Quilodran, R., Rothé, M., & Procyk, E. (2008). Behavioral Shifts and Action Valuation in the Anterior Cingulate Cortex. *Neuron*, *57*(2), 314–325. https://doi.org/10.1016/j.neuron.2007.11.031

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, *19*(1), 165–171. https://doi.org/10.1038/nn.4179

Rubinov, M., & Sporns, O. (2010). NeuroImage Complex network measures of brain connectivity : Uses and interpretations. *NeuroImage*, *52*(3), 1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

Rudie, J. D., Brown, J. A., Beck-Pancer, D., Hernandez, L. M., Dennis, E. L., Thompson, P. M., … Dapretto, M. (2013). Altered functional and structural brain network organization in autism. *NeuroImage: Clinical*, *2*(1), 79–94. https://doi.org/10.1016/j.nicl.2012.11.006

Rypma, B., Berger, J. S., Prabhakaran, V., Bly, B. M., Kimberg, D. Y., Biswal, B. B., & D'Esposito, M. (2006). Neural correlates of cognitive efficiency. *NeuroImage*, *33*(3), 969–979. https://doi.org/10.1016/j.neuroimage.2006.05.065

Sadaghiani, S., & D'Esposito, M. (2015). Functional characterization of the cingulo-opercular network in the maintenance of tonic alertness. *Cerebral Cortex*, *25*(9), 2763–2773. https://doi.org/10.1093/cercor/bhu072

Sadaghiani, S., & Kleinschmidt, A. (2016). Brain Networks and α-Oscillations: Structural and Functional Foundations of Cognitive Control. *Trends in Cognitive Sciences*, *20*(11), 805–817. https://doi.org/10.1016/j.tics.2016.09.004

Sadaghiani, S., Poline, J.-B., Kleinschmidt, A., & D'Esposito, M. (2015). Ongoing dynamics in large-scale functional connectivity predict perception. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(27), 8463–8468. https://doi.org/10.1073/pnas.1420687112

Sadaghiani, S., Scheeringa, R., Lehongre, K., Morillon, B., Giraud, A.-L., & Kleinschmidt, A. (2010). Intrinsic Connectivity Networks, Alpha Oscillations, and Tonic Alertness: A Simultaneous Electroencephalography/Functional Magnetic Resonance Imaging Study.

Journal of Neuroscience, *30*(30), 10243–10250. https://doi.org/10.1523/JNEUROSCI.1004-10.2010

Sakai, K. (2008). Task set and prefrontal cortex. *Annual Review of Neuroscience*, *31*, 219–245. https://doi.org/10.1146/annurev.neuro.31.060407.125642

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403–428. https://doi.org/10.1037/0033-295x.103.3.403

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., … Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*, 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., … Yeo, B. T. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179

Seeley, W. W. (2019). The Salience Network: A Neural System for Perceiving and Responding to Homeostatic Demands. *The Journal of Neuroscience*, *39*(50), 9878–9882. https://doi.org/10.1523/JNEUROSCI.1138-17.2019

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., … Greicius, M. D. (2007). Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. *Journal of Neuroscience*, *27*(9), 2349–2356. https://doi.org/10.1523/JNEUROSCI.5587-06.2007

Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, *5*(September), 1–17. https://doi.org/10.3389/fpubh.2017.00258

Sherwin, J., Muraskin, J., & Sajda, P. (2012). You Can't Think and Hit at the Same Time: Neural Correlates of Baseball Pitch Classification. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00177

Shine, J. M., Bissett, P. G., Bell, P. T., Koyejo, O., Balsters, J. H., Gorgolewski, K. J., … Poldrack, R. A. (2016). The Dynamics of Functional Brain Networks: Integrated Network States during Cognitive Task Performance. *Neuron*, *92*(2), 544–554. https://doi.org/10.1016/j.neuron.2016.09.018

Shine, J. M., & Breakspear, M. (2018). Understanding the Brain, By Default. *Trends in Neurosciences*, *41*(5), 244–247. https://doi.org/10.1016/j.tins.2018.03.004

Sie, J.-H., Chen, Y.-H., Chang, C.-Y., Yen, N.-S., Chu, W.-C., & Shiau, Y.-H. (2019). Altered Central Autonomic Network in Baseball Players: A Resting-state fMRI Study. *Scientific Reports*, *9*(1), 110. https://doi.org/10.1038/s41598-018-36329-9

Smith, A. C., Frank, L. M., Wirth, S., Yanike, M., Hu, D., Kubota, Y., … Brown, E. N. (2004). Dynamic Analysis of Learning in Behavioral Experiments. *J. Neurosci.*, *24*(2), 447–461. https://doi.org/10.1523/JNEUROSCI.2908-03.2004

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., … Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences*, *106*(31), 13040–13045. https://doi.org/10.1073/pnas.0905267106

Smith, Stephen M., Nichols, T. E., Vidaurre, D., Winkler, A. M., Behrens, T. E. J., Glasser, M. F., … Miller, K. L. (2015). A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience*, *18*(11), 1565–1567. https://doi.org/10.1038/nn.4125

Sporns, O., & Betzel, R. F. (2016). Modular Brain Networks. *Annual Review of Psychology*, *67*(1), 613–640. https://doi.org/10.1146/annurev-psych-122414-033634

Spreng, R. N., Stevens, W. D., Chamberlain, J. P., Gilmore, A. W., & Schacter, D. L. (2010). Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *NeuroImage*, *53*(1), 303–317. https://doi.org/10.1016/j.neuroimage.2010.06.016

Stevens, A. A., Tappon, S. C., Garg, A., & Fair, D. A. (2012). Functional Brain Network Modularity Captures Inter- and Intra-Individual Variation in Working Memory Capacity, *7*(1). https://doi.org/10.1371/journal.pone.0030468

Stevner, A. B. A., Vidaurre, D., Cabral, J., Rapuano, K., Nielsen, S. F. V., Tagliazucchi, E., … Kringelbach, M. L. (2019). Discovery of key whole-brain transitions and dynamics during human wakefulness and non-REM sleep. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-08934-3

Stoll, F. M., Fontanier, V., & Procyk, E. (2016). Specific frontal neural dynamics contribute to decisions to check. *Nature Communications*, *7*(May). https://doi.org/10.1038/ncomms11990

Strack, B. W. (2003). Effect of heart rate variability (hrv) biofeedback on batting performance in baseball. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *64*(3-B).

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Thayer, J. F., Hansen, A. L., Saus-Rose, E., & Johnsen, B. H. (2009). Heart rate variability, prefrontal neural function, and cognitive performance: The neurovisceral integration perspective on self-regulation, adaptation, and health. *Annals of Behavioral Medicine*, *37*(2), 141–153. https://doi.org/10.1007/s12160-009-9101-z

Thompson, G. J. (2018). Neural and metabolic basis of dynamic resting state fMRI. *NeuroImage*, *180*, 448–462. https://doi.org/10.1016/j.neuroimage.2017.09.010

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

Vatansever, D., Menon, D. K., Manktelow, A. E., Sahakian, B. J., & Stamatakis, E. A. (2015). Default Mode Dynamics for Global Functional Integration. *Journal of Neuroscience*, *35*(46), 15254–15262. https://doi.org/10.1523/JNEUROSCI.2135-15.2015

Vidaurre, D., Abeysuriya, R., Becker, R., Quinn, A. J., Alfaro-Almagro, F., Smith, S. M., & Woolrich, M. W. (2018). Discovering dynamic brain networks from big data in rest and task. *NeuroImage*, *180*, 646–656. https://doi.org/10.1016/j.neuroimage.2017.06.077

Vidaurre, D., Smith, S. M., & Woolrich, M. W. (2017). Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences*, *114*(48), 201705120. https://doi.org/10.1073/pnas.1705120114

Vlemincx, E., Taelman, J., De Peuter, S., Van Diest, I., & Van Den Bergh, O. (2011). Sigh rate

and respiratory variability during mental load and sustained attention. *Psychophysiology*, *48*(1), 117–120. https://doi.org/10.1111/j.1469-8986.2010.01043.x

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *4*(3), 274–290. https://doi.org/10.1111/j.1745-6924.2009.01125.x

Wainwright, M. J., & Jordan, M. I. (2007). Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends® in Machine Learning*, *1*(1–2), 1–305. https://doi.org/10.1561/2200000001

Walton, M. E., Bannerman, D. M., Alterescu, K., & Rushworth, M. F. S. (2003). Functional specialization within medial frontal cortex of the anterior cingulate for evaluating effort-related decisions. *Journal of Neuroscience*, *23*(16), 6475–6479. https://doi.org/10.1523/jneurosci.23-16-06475.2003

Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., … Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience*, *21*(6), 860–868. https://doi.org/10.1038/s41593-018-0147-8

Woodworth, R. S., & Thorndike, E. L. (1901). The influence of improvement in one mental function upon the efficiency of other functions. (I). *Psychological Review*, *8*(3), 247–261. https://doi.org/10.1037/h0074898

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., … Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424

Zuo, X.-N., He, Y., Betzel, R. F., Colcombe, S., Sporns, O., & Milham, M. P. (2017). Human Connectomics across the Life Span. *Trends in Cognitive Sciences*, *21*(1), 32–45. https://doi.org/10.1016/j.tics.2016.10.005